```
Db 150515 CCTCTCCCCAGGATTCTTGGAACAGGGCATCCTGGTGGTGGACAAGGGTAGGATCTCGAG 150574

Qy    321 tcgctacgttcgcacctggagtttcttcttggacctggcttccctgatgcccacagatgt 380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 150575 TCGCTACGTTCGCACCTGGAGTTTCTTCTTGGACCTGGCTTCCCTGATGCCCACAGATGT 150634

Qy    381 ggtctacgtgcggctgggcccgcacacacccaccctgaggctgaaccgctttctccgcgc 440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 150635 GGTCTACGTGCGGCTGGGCCCGCACACACCCACCCTGAGGCTGAACCGCTTTCTCCGCGC 150694

Qy    441 gccccgcctcttcgaggccttcgaccgcacagagacccgcacagcttacccaaatgcctt 500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 150695 GCCCCGCCTCTTCGAGGCCTTCGACCGCACAGAGACCCGCACAGCTTACCCAAATGCCTT 150754

Qy    501 tcgcattgccaagctgatgctttacatttttgtcgtcatccattggaacagctgcctata 560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 150755 TCGCATTGCCAAGCTGATGCTTTACATTTTTGTCGTCATCCATTGGAACAGCTGCCTATA 150814

Qy    561 ctttgccctatcccggtacctgggcttcgggcgtgacgcatgggtgtacccggaccccgc 620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 150815 CTTTGCCCTATCCCGGTACCTGGGCTTCGGGCGTGACGCATGGGTGTACCCGGACCCCGC 150874

Qy    621 gcagcctggctttgagcgccctgcggcgccagtacctctatagcttttacttctccacgct 680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 150875 GCAGCCTGGCTTTGAGCGCCTGCGGCGCCAGTACCTCTATAGCTTTTACTTCTCCACGCT 150934

Qy    681 gatactgactacagtgggcgatacaccgccgccagccagggaagaagagtacctcttcat 740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 150935 GATACTGACTACAGTGGGCGATACACCGCCGCCAGCCAGGGAAGAAGAGTACCTCTTCAT 150994

Qy    741 ggtgggcgacttcctgctggccgtcatgggtttcgccaccatcatgggtagcatgagctc 800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 150995 GGTGGGCGACTTCCTGCTGGCCGTCATGGGTTTCGCCACCATCATGGGTAGCATGAGCTC 151054

Qy    801 tgtcatctacaacatgaacactgcagatgcggctttctacccagatcatgcactggtgaa 860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 151055 TGTCATCTACAACATGAACACTGCAGATGCGGCTTTCTACCCAGATCATGCACTGGTGAA 151114

Qy    861 gaagtacatgaagctgcagcacgtcaaccgcaagctggagcggcgagttattgactggt 919
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 151115 GAAGTACATGAAGCTGCAGCACGTCAACCGCAAGCTGGAGCGGCGAGTTATTGACTGGT 151173
```

```
RESULT   7
AC036216
LOCUS       AC036216              164434 bp    DNA     linear   HTG 04-MAY-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-658A5 map 11, WORKING DRAFT
            SEQUENCE, 26 unordered pieces.
ACCESSION   AC036216
VERSION     AC036216.2  GI:7705203
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 164434)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-658A5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 164434)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Labocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 4, 2000 this sequence version replaced gi:7523886.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L9404
                Center clone name: 658_A_5
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 149016 bases at least Q40
                Consensus quality: 157301 bases at least Q30
                Consensus quality: 160298 bases at least Q20
                Insert size: 157000; agarose-fp
                Insert size: 161934; sum-of-contigs
                Quality coverage: 3.7 in Q20 bases; agarose-fp
                Quality coverage: 3.6 in Q20 bases; sum-of-contigs
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1416: contig of 1416 bp in length
            *     1417  1516: gap of      100 bp
            *     1517     3580: contig of 2064 bp in length
            *     3581  3680: gap of      100 bp
            *     3681     4383: contig of 703 bp in length
            *     4384  4483: gap of      100 bp
            *     4484     5844: contig of 1361 bp in length
            *     5845  5944: gap of      100 bp
            *     5945     7574: contig of 1630 bp in length
            *     7575  7674: gap of      100 bp
            *     7675     9988: contig of 2314 bp in length
            *     9989 10088: gap of      100 bp
            *    10089    13426: contig of 3338 bp in length
            *    13427 13526: gap of      100 bp
            *    13527    15573: contig of 2047 bp in length
            *    15574 15673: gap of      100 bp
            *    15674    18974: contig of 3301 bp in length
            *    18975 19074: gap of      100 bp
            *    19075    21809: contig of 2735 bp in length
            *    21810 21909: gap of      100 bp
            *    21910    25085: contig of 3176 bp in length
            *    25086 25185: gap of      100 bp
            *    25186    29425: contig of 4240 bp in length
            *    29426 29525: gap of      100 bp
            *    29526    34133: contig of 4608 bp in length
            *    34134 34233: gap of      100 bp
            *    34234    40649: contig of 6416 bp in length
            *    40650 40749: gap of      100 bp
            *    40750    45891: contig of 5142 bp in length
            *    45892 45991: gap of      100 bp
```

THIS PAGE BLANK (USPTO)

```
            *    45992    53180: contig of 7189 bp in length
            *    53181    53280: gap of      100 bp
            *    53281    62817: contig of 9537 bp in length
            *    62818    62917: gap of      100 bp
            *    62918    70247: contig of 7330 bp in length
            *    70248    70347: gap of      100 bp
            *    70348    78754: contig of 8407 bp in length
            *    78755    78854: gap of      100 bp
            *    78855    86840: contig of 7986 bp in length
            *    86841    86940: gap of      100 bp
            *    86941    98610: contig of 11670 bp in length
            *    98611    98710: gap of      100 bp
            *    98711   111069: contig of 12359 bp in length
            *   111070   111169: gap of      100 bp
            *   111170   122730: contig of 11561 bp in length
            *   122731   122830: gap of      100 bp
            *   122831   135722: contig of 12892 bp in length
            *   135723   135822: gap of      100 bp
            *   135823   150963: contig of 15141 bp in length
            *   150964   151063: gap of      100 bp
            *   151064   164434: contig of 13371 bp in length.
FEATURES             Location/Qualifiers
     source          1..164434
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-658A5"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..1416
                     /note="assembly_fragment"
     misc_feature    1517..3580
                     /note="assembly_fragment"
     misc_feature    3681..4383
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    4484..5844
                     /note="assembly_fragment"
     misc_feature    5945..7574
                     /note="assembly_fragment"
     misc_feature    7675..9988
                     /note="assembly_fragment"
     misc_feature    10089..13426
                     /note="assembly_fragment"
     misc_feature    13527..15573
                     /note="assembly_fragment"
     misc_feature    15674..18974
                     /note="assembly_fragment"
     misc_feature    19075..21809
                     /note="assembly_fragment"
     misc_feature    21910..25085
                     /note="assembly_fragment"
     misc_feature    25186..29425
                     /note="assembly_fragment"
     misc_feature    29526..34133
                     /note="assembly_fragment"
     misc_feature    34234..40649
                     /note="assembly_fragment"
     misc_feature    40750..45891
                     /note="assembly_fragment"
     misc_feature    45992..53180
                     /note="assembly_fragment"
     misc_feature    53281..62817
                     /note="assembly_fragment"
     misc_feature    62918..70247
                     /note="assembly_fragment"
     misc_feature    70348..78754
                     /note="assembly_fragment"
     misc_feature    78855..86840
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    86941..98610
                     /note="assembly_fragment"
     misc_feature    98711..111069
                     /note="assembly_fragment"
     misc_feature    111170..122730
                     /note="assembly_fragment"
     misc_feature    122831..135722
                     /note="assembly_fragment"
     misc_feature    135823..150963
                     /note="assembly_fragment"
     misc_feature    151064..164434
                     /note="assembly_fragment"
BASE COUNT    50098 a   34031 c   34547 g   43240 t    2518 others
ORIGIN


  Query Match             37.8%;  Score 652.6;  DB 2;  Length 164434;
  Best Local Similarity   99.4%;  Pred. No. 6.8e-142;
  Matches  655;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      261 cttccacacaggattcttggaacagggcatcctggtggtggacaagggtaggatctcgag 320
            | ||  | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   157488 CCTCTCCCCAGGATTCTTGGAACAGGGCATCCTGGTGGTGGACAAGGGTAGGATCTCGAG 157547

Qy      321 tcgctacgttcgcacctggagtttcttcttggacctggcttccctgatgcccacagatgt 380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   157548 TCGCTACGTTCGCACCTGGAGTTTCTTCTTGGACCTGGCTTCCCTGATGCCCACAGATGT 157607

Qy      381 ggtctacgtgcggctgggcccgcacacacccaccctgaggctgaaccgctttctccgcgc 440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   157608 GGTCTACGTGCGGCTGGGCCCGCACACACCCACCCTGAGGCTGAACCGCTTTCTCCGCGC 157667

Qy      441 gccccgcctcttcgaggccttcgaccgcacagagacccgcacagcttacccaaatgcctt 500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   157668 GCCCCGCCTCTTCGAGGCCTTCGACCGCACAGAGACCCGCACAGCTTACCCAAATGCCTT 157727

Qy      501 tcgcattgccaagctgatgctttacatttttgtcgtcatccattggaacagctgcctata 560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   157728 TCGCATTGCCAAGCTGATGCTTTACATTTTTGTCGTCATCCATTGGAACAGCTGCCTATA 157787

Qy      561 ctttgccctatcccggtacctgggcttcgggcgtgacgcatgggtgtacccggaccccgc 620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   157788 CTTTGCCCTATCCCGGTACCTGGGCTTCGGGCGTGACGCATGGGTGTACCCGGACCCCGC 157847

Qy      621 gcagcctggctttgagcgcctgcggcgccagtacctctatagcttttacttctccacgct 680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   157848 GCAGCCTGGCTTTGAGCGCCTGCGGCGCCAGTACCTCTATAGCTTTTACTTCTCCACGCT 157907

Qy      681 gatactgactacagtgggcgatacaccgccgccagccagggaagaagagtacctcttcat 740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   157908 GATACTGACTACAGTGGGCGATACACCGCCGCCAGCCAGGGAAGAAGAGTACCTCTTCAT 157967

Qy      741 ggtgggcgacttcctgctggccgtcatgggtttcgccaccatcatgggtagcatgagctc 800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   157968 GGTGGGCGACTTCCTGCTGGCCGTCATGGGTTTCGCCACCATCATGGGTAGCATGAGCTC 158027

Qy      801 tgtcatctacaacatgaacactgcagatgcggctttctacccagatcatgcactggtgaa 860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   158028 TGTCATCTACAACATGAACACTGCAGATGCGGCTTTCTACCCAGATCATGCACTGGTGAA 158087

Qy      861 gaagtacatgaagctgcagcacgtcaaccgcaagctggagcggcgagttattgactggt 919
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   158088 GAAGTACATGAAGCTGCAGCACGTCAACCGCAAGCTGGAGCGGCGAGTTATTGACTGGT 158146


RESULT   8
AC022762
LOCUS       AC022762              186609 bp    DNA     linear   HTG 09-FEB-2002
DEFINITION  Homo sapiens chromosome 11 clone RP11-290F24 map 11, *** SEQUENCING
            IN PROGRESS ***, 1 ordered pieces.
ACCESSION   AC022762
VERSION     AC022762.6  GI:18642872
KEYWORDS    HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
```

THIS PAGE BLANK (USPTO)

OM of: US-09-927-267-1 to: N_Geneseq_032802:* out_format : pfs

Date: Jun 20, 2002 6:35 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09927267/runat_19062002_090143_25524/app_query.fasta_1.640
-DB=N_Geneseq_032802 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09927267_SCGN1_1_663 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-927-267-1
Query length: 575
Database: N_Geneseq_032802:*
Database sequences: 1736436
Database length: 858457221
Search time (sec): 267.690000

score_list:

Sequence	Strd	Orig	%Score	EScore	Len	! Documentation ..		
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:ABA04603	+	2984.00	5715.50	0.0				25
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:ABA04602	+	2957.50	5668.28	0.0				18
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAD09558	+	1535.00	2923.86	1.6e-154				1
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL13353	+	1235.00	2340.84	4.9e-122				2
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH98838	+	959.50	1822.14	3.8e-93				9
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL18085	+	947.00	1789.19	2.6e-91				2
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL14931	+	867.50	1625.39	3.5e-82				5
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL13352	+	863.50	1611.61	2.0e-81				8
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL19457	+	781.50	1464.82	3.0e-73				3
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS91658	+	714.00	1332.86	6.8e-66				3
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL19456	-	695.00	1291.49	1.4e-63				5
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:AAD25729	+	655.00	1222.88	9.1e-60				27
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS91258	+	571.00	1071.44	2.5e-51				1
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS79945	+	540.00	1011.43	5.5e-48				1
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS77668	+	540.00	1002.62	1.7e-47				2
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL06311	+	503.00	938.76	6.1e-44				1
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL18084	-	502.00	906.28	3.9e-42				1
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAZ09486	+	480.50	884.64	6.3e-41				31
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL14930	-	478.00	863.46	9.5e-40				1
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAZ09496	+	467.50	854.86	2.9e-39				47
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH48730	+	467.50	854.10	3.2e-39				5
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:ABA09197	+	467.50	853.14	3.6e-39				5
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAX84444	+	466.50	865.87	7.0e-40				15
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH48729	+	465.50	855.04	2.8e-39				3
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH48731	+	464.50	854.09	3.2e-39				3
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC66779	+	456.50	842.06	1.5e-38				23
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAZ09488	+	451.00	834.34	4.0e-38				18
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAZ09485	+	445.00	816.94	3.7e-37				29
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:AAS18223	+	441.50	815.68	4.4e-37				18
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:AAS18224	+	441.50	815.68	4.4e-37				18
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAX84443	+	441.50	813.45	5.8e-37				22
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAX84442	+	441.50	811.23	7.7e-37				27
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:AAS18213	+	441.50	810.98	8.0e-37				27
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:AAS18215	+	441.50	810.98	8.0e-37				27
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:AAS18214	+	441.50	810.21	8.8e-37				29
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:AAS18216	+	441.50	810.21	8.8e-37				29
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAZ09493	+	441.50	810.19	8.8e-37				29
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAZ09484	+	435.50	800.22	3.2e-36				26
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAZ09491	+	432.00	792.26	8.8e-36				28
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAZ09487	+	430.50	789.23	1.3e-35				29
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL14239	+	422.00	769.36	1.7e-34				
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAX84445	+	403.50	742.93	4.9e-33				
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH98302	+	400.50	729.48	2.8e-32				
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAA62050	+	400.50	729.45	2.8e-32				
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAZ35716	+	395.50	721.87	7.3e-32				

seq_name: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:ABA04603

seq_documentation_block:
ID ABA04603 standard; cDNA; 2551 BP.
XX
AC ABA04603;
XX
DT 21-FEB-2002 (first entry)
XX
DE MOL10b coding sequence.
XX
KW MOL; G-coupled protein-receptor; cardiomyopathy; atherosclerosis;
KW cell signal processing; metabolic disorder; diabetes; cancer; human;
KW neurodegenerative disorder; immune disorder; cardiac disorder;
KW lung disease; autoimmune disease; developmental disorder; antidiabetic;
KW Cytostatic; Neuroprotective; Antiatherosclerotic; Immunosuppressive;
KW Gene therapy; Vaccine; antiinflammatory; MOL10b; chromosome 11; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 779..2506
FT /*tag= a
FT /product= "MOL 10b"
XX
PN WO200181578-A2.
XX
PD 01-NOV-2001.
XX
PF 26-APR-2001; 2001WO-US13578.
XX
PR 26-APR-2000; 2000US-200158P.
PR 28-APR-2000; 2000US-200613P.
PR 28-APR-2000; 2000US-200780P.
PR 01-MAY-2000; 2000US-201006P.
PR 01-MAY-2000; 2000US-201007P.
PR 01-MAY-2000; 2000US-201236P.
PR 01-MAY-2000; 2000US-201238P.
PR 02-MAY-2000; 2000US-201186P.
PR 03-MAY-2000; 2000US-201474P.
PR 03-MAY-2000; 2000US-201508P.
PR 25-JUL-2000; 2000US-220591P.
PR 15-SEP-2000; 2000US-232678P.
PR 22-JAN-2001; 2001US-263217P.
PR 30-JAN-2001; 2001US-265160P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Vernet CAM, Fernandes ER, Gerlach V, Shimkets RA, Malyankar UM;
PI Boldog FL, Zerhusen BD, Spytek KA, Majumder K, Tchernev VT;
PI Padigaru M, Patturajan M, Burgess CE, Gangolli EA, Smithson G;
PI Rastelli L, Macdougall JR, Taupier RJ, Grosse WM, Szekeres ES;
PI Alsobrook JP;
XX
DR WPI; 2002-049278/06.
DR P-PSDB; AAM47673.
XX
PT Novel G-protein coupled receptor-related polypeptides and
PT polynucleotides for diagnosing, preventing and treating cardiomyopathy,
PT atherosclerosis, disorders related to cell signal processing and for
PT identifying modulators -
XX
PS Claim 8; Page 78; 227pp; English.
XX
CC The present invention relates to novel G-coupled protein-receptor related
CC proteins and coding sequences (MOLX, where X is a number from 1 to 10,
CC ABA04589-ABA04603 and AAM47659-AAM47673). MOLX proteins and coding

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

CC sequences are useful for treating or preventing a MOLX-associated
CC disorder, such as cardiomyopathy, atherosclerosis, disorders related to
CC cell signal processing and metabolic pathway modulation, diabetes and
CC cancer. Additionally, MOLX proteins and coding sequences are useful for
CC preventing and treating a variety of disorders including metabolic
CC disorders, nutritional oedema, chronic and hereditary pancreatitis,
CC obesity, infectious disease, anorexia, neurodegenerative disorders,
CC Alzheimer's disease, Parkinson's disease, stroke, immune disorders,
CC haematopoietic disorders and various dyslipidaemias, metabolic syndrome X
CC and wasting disorders associated with chronic diseases and cancers,
CC cardiac disorders, hypertension, hypercalcaemia, cirrhosis, angiogenesis
CC and wound healing, trauma, glomerulonephritis, hyper and hypothyroidism,
CC multiple sclerosis, lung diseases including asthma, Crohn's disease,
CC scleroderma, autoimmune diseases, developmental disorders and neural tube
CC defects. The present sequence is the coding sequence for human MOL10b,
CC which maps to chromosome 11. MOL10b protein is a cyclic nucleotide-gated
CC olfactory channel-like protein.
XX
Sequence 2551 BP; 646 A; 668 C; 707 G; 530 T; 0 other;

(see Table 10E, page 79)

102(a)

alignment_scores:
Quality: 2984.00 Length: 575
Ratio: 5.190 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.826

alignment_block:
US-09-927-267-1 x ABA04603 ..

Align seg 1/1 to: ABA04603 from: 1 to: 2551

```
      1 MetSerGlnAspThrLysValLysThrThrGluSerSerProProAlaPr 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    779 ATGAGCCAGGACACCAAAGTGAAGACAACAGAGTCCAGTCCCCCAGCCCC 828

     17 oSerLysAlaArgLysLeuLeuProValLeuAspProSerGlyAspTyrT 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    829 ATCCAAGGCCAGGAAGTTGCTGCCTGTCCTGGACCCATCTGGGGATTACT 878

     34 yrTyrTrpTrpLeuAsnThrMetValProProValMetTyrAsnLeuIle 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    879 ACTACTGGTGGCTGAACACAATGGTCTTCCCAGTCATGTATAACCTCATC 928

     51 IleLeuValCysArgAlaCysProProAspLeuGlnHisGlyTyrLeuVa 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    929 ATCCTCGTGTGCAGAGCCTGCTTCCCCGACTTGCAGCACGGTTATCTGGT 978

     67 lAlaTrpLeuValLeuAspTyrThrSerAspLeuLeuTyrLeuLeuAspM 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    979 GGCCTGGTTGGTGCTGGACTACACGAGTGACCTGCTATACCTACTAGACA 1028

     84 etValValArgPheHisThrGlyPheLeuGluGlnGlyIleLeuValVal 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1029 TGGTGGTGCGCTTCCACACAGGATTCTTGGAACAGGGCATCCTGGTGGTG 1078

    101 AspLysGlyArgIleSerSerArgTyrValArgThrTrpSerPhePheLe 117
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1079 GACAAGGGTAGGATCTCGAGTCGCTACGTTCGCACCTGGAGTTTCTTCTT 1128

    117 uAspLeuAlaSerLeuMetProThrAspValValTyrValArgLeuGlyP 134
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1129 GGACCTGGCTTCCCTGATGCCCACAGATGTGGTCTACGTGCGGCTGGGCC 1178

    134 roHisThrProThrLeuArgLeuAsnArgPheLeuArgAlaProArgLeu 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1179 CGCACACACCCACCCTGAGGCTGAACCGCTTTCTCCGCGCGCCCCGCCTC 1228

                 157 T→A
    151 PheGluAlaPheAspArgThrGluThrArgThrAlaTyrProAsnAlaPh 167
        ||||||||||||||||||::||||||||||||||||||||||||||||||
   1229 TTCGAGGCCTTCGACCGCGCAGAGACCCGCACAGCTTACCCAAATGCCTT 1278

    167 eArgIleAlaLysLeuMetLeuTyrIlePheValValIleHisTrpAsnS 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
   1279 TCGCATTGCCAAGCTGATGCTTTACATTTTGTCGTCATCCATTGGAACA 1328

    184 erCysLeuTyrPheAlaLeuSerArgTyrLeuGlyPheGlyArgAspAla 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1329 GCTGCCTATACTTTGCCCTATCCCGGTACCTGGGCTTCGGGCGTGACGCA 1378

    201 TrpValTyrProAspProAlaGlnProGlyPheGluArgLeuArgArgGl 217
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1379 TGGGTGTACCCGGACCCCGCGCAGCCTGGCTTTGAGCGCCTGCGGCGCCA 1428

    217 nTyrLeuTyrSerPheTyrPheSerThrLeuIleLeuThrThrValGlyA 234
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1429 GTACCTCTATAGCTTTTACTTCTCCACGCTGATACTGACTACAGTGGGCG 1478

    234 spThrProProProAlaArgGluGluGluTyrLeuPheMetValGlyAsp 250
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1479 ATACACCGCCGCCAGCCAGGGAAGAAGAGTACCTCTTCATGGTGGGCGAC 1528

    251 PheLeuLeuAlaValMetGlyPheAlaThrIleMetGlySerMetSerSe 267
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1529 TTCCTGCTGGCCGTCATGGGTTTCGCCACCATCATGGGTAGCATGAGCTC 1578

    267 rValIleTyrAsnMetAsnThrAlaAspAlaAlaPheTyrProAspHisA 284
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1579 TGTCATCTACAACATGAACACTGCAGATGCGGCTTTCTACCCAGATCATG 1628

    284 laLeuValLysLysTyrMetLysLeuGlnHisValAsnArgLysLeuGlu 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1629 CACTGGTGAAGAAGTACATGAAGCTGCAGCACGTCAACCGCAAGCTGGAG 1678

    301 ArgArgValIleAspTrpTyrGlnHisLeuGlnIleAsnLysLysMetTh 317
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1679 CGGCGAGTTATTGACTGGTATCAGCACCTGCAGATCAACAAGAAGATGAC 1728

    317 rAsnGluValAlaIleLeuGlnHisLeuProGluArgLeuArgAlaGluV 334
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1729 CAACGAGGTAGCCATCTTACAGCACTTGCCTGAGCGGCTGCGGGCAGAAG 1778

    334 alAlaValSerValHisLeuSerThrLeuSerArgValGlnIlePheGln 350
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1779 TGGCTGTGTCTGTGCACCTGTCCACTCTGAGCCGGGTGCAGATCTTTCAG 1828

    351 AsnCysGluAlaSerLeuLeuGluGluLeuValLeuLysLeuGlnProGl 367
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1829 AACTGTGAGGCCAGCCTGCTGGAGGAGCTGGTGCTGAAGCTGCAGCCCCA 1878

    367 nThrTyrSerProGlyGluTyrValCysArgLysGlyAspIleGlyGlnG 384
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1879 GACCTACTCACCAGGTGAATATGTATGCCGCAAAGGAGACATTGGCCAAG 1928

    384 luMetTyrIleIleArgGluGlyGlnLeuAlaValValAlaAspAspGly 400
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1929 AGATGTACATCATCCGAGAGGGTCAACTGGCCGTGGTGGCAGATGATGGT 1978

    401 IleThrGlnTyrAlaValLeuGlyAlaGlyLeuTyrPheGlyGluIleSe 417
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1979 ATCACACAGTATGCTGTGCTCGGTGCAGGGCTCTACTTTGGGGAGATCAG 2028

    417 rIleIleAsnIleLysGlyAsnMetSerGlyAsnArgArgThrAlaAsnI 434
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2029 CATCATCAACATCAAAGGGAACATGTCTGGGAACCGCCGCACAGCCAACA 2078

    434 leLysSerLeuGlyTyrSerAspLeuPheCysLeuSerLysGluAspLeu 450
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2079 TCAAGAGCCTAGGTTATTCAGACCTATTCTGCCTGAGCAAGGAGGACCTG 2128

    451 ArgGluValLeuSerGluTyrProGlnAlaGlnThrIleMetGluGluLy 467
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2129 CGGGAGGTGCTGAGCGAGTATCCACAAGCACAGACCATCATGGAGGAGAA 2178

    467 sGlyArgGluIleLeuLeuLysMetAsnLysLeuAspValAsnAlaGluA 484
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2179 AGGACGTGAGATCCTGCTGAAAATGAACAAGTTGGACGTGAATGCTGAGG 2228
```

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

```
 484 laAlaGluIleAlaLeuGlnGluAlaThrGluSerArgLeuArgGlyLeu 500
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2229 CAGCTGAGATCGCCCTGCAGGAGGCCACAGAGTCCCGGCTACGAGGCCTA 2278

 501 AspGlnGlnLeuAspAspLeuGlnThrLysPheAlaArgLeuLeuAlaGl 517
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2279 GACCAGCAGCTGGATGATCTACAGACCAAGTTTGCTCGCCTCCTGGCTGA 2328

 517 uLeuGluSerSerAlaLeuLysIleAlaTyrArgIleGluArgLeuGluT 534
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2329 GCTGGAGTCCAGCGCACTTAAGATTGCTTACCGCATTGAACGGCTGGAGT 2378

 534 rpGlnThrArgGluTrpProMetProGluAspLeuAlaGluAlaAspAsp 550
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2379 GGCAGACTCGAGAGTGGCCAATGCCCGAGGACCTGGCTGAGGCTGATGAC 2428

 551 GluGlyGluProGluGluGlyThrSerLysAspGluGluGlyArgAlaSe 567
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2429 GAGGGTGAGCCTGAGGAGGGAACTTCCAAAGATGAAGAGGGCAGGGCCAG 2478

 567 rGlnGluGlyProProGlyProGlu 575
     |||||||||||||||||||||||||
2479 CCAGGAGGGACCCCCAGGTCCAGAG 2503
```

seq_name: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:ABA04602

seq_documentation_block:

```
ID   ABA04602 standard; cDNA; 1835 BP.
XX
AC   ABA04602;
XX
DT   21-FEB-2002  (first entry)
XX
DE   MOL10a coding sequence.
XX
KW   MOL; G-coupled protein-receptor; cardiomyopathy; atherosclerosis;
KW   cell signal processing; metabolic disorder; diabetes; cancer;
KW   neurodegenerative disorder; immune disorder; cardiac disorder;
KW   lung disease; autoimmune disease; developmental disorder; antidiabetic;
KW   Cytostatic; Neuroprotective; Antiatherosclerotic; Immunosuppressive;
KW   Gene therapy; Vaccine; antiinflammatory; MOL10a; ss.
XX
OS   Unidentified.
XX
FH   Key             Location/Qualifiers
FT   CDS             54..1790
FT                   /*tag= a
FT                   /product= "MOL 10a"
XX
PN   WO200181578-A2.
XX
PD   01-NOV-2001.
XX
PF   26-APR-2001; 2001WO-US13578.
XX
PR   26-APR-2000; 2000US-200158P.
PR   28-APR-2000; 2000US-200613P.
PR   28-APR-2000; 2000US-200780P.
PR   01-MAY-2000; 2000US-201006P.
PR   01-MAY-2000; 2000US-201007P.
PR   01-MAY-2000; 2000US-201236P.
PR   01-MAY-2000; 2000US-201238P.
PR   02-MAY-2000; 2000US-201186P.
PR   03-MAY-2000; 2000US-201474P.
PR   03-MAY-2000; 2000US-201508P.
PR   25-JUL-2000; 2000US-220591P.
PR   15-SEP-2000; 2000US-232678P.
PR   22-JAN-2001; 2001US-263217P.
PR   30-JAN-2001; 2001US-265160P.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Vernet CAM,  Fernandes ER,  Gerlach V,  Shimkets RA,  Malyankar UM;
PI   Boldog FL,  Zerhusen BD,  Spytek KA,  Majumder K,  Tchernev VT;
PI   Padigaru M,  Patturajan M,  Burgess CE,  Gangolli EA,  Smithson G;
PI   Rastelli L,  Macdougall JR,  Taupier RJ,  Grosse WM,  Szekeres ES;
PI   Alsobrook JP;
XX
DR   WPI; 2002-049278/06.
DR   P-PSDB; AAM47672.
XX
PT   Novel G-protein coupled receptor-related polypeptides and
PT   polynucleotides for diagnosing, preventing and treating cardiomyopathy,
PT   atherosclerosis, disorders related to cell signal processing and for
PT   identifying modulators  -
XX
PS   Claim 8; Page 76; 227pp; English.
XX
CC   The present invention relates to novel G-coupled protein-receptor related
CC   proteins and coding sequences (MOLX, where X is a number from 1 to 10,
CC   ABA04589-ABA04603 and AAM47659-AAM47673). MOLX proteins and coding
CC   sequences are useful for treating or preventing a MOLX-associated
CC   disorder, such as cardiomyopathy, atherosclerosis, disorders related to
CC   cell signal processing and metabolic pathway modulation, diabetes and
CC   cancer. Additionally, MOLX proteins and coding sequences are useful for
CC   preventing and treating a variety of disorders including metabolic
CC   disorders, nutritional oedema, chronic and hereditary pancreatitis,
CC   obesity, infectious disease, anorexia, neurodegenerative disorders,
CC   Alzheimer's disease, Parkinson's disease, stroke, immune disorders,
CC   haematopoietic disorders and various dyslipidaemias, metabolic syndrome X
CC   and wasting disorders associated with chronic diseases and cancers,
CC   cardiac disorders, hypertension, hypercalcaemia, cirrhosis, angiogenesis
CC   and wound healing, trauma, glomerulonephritis, hyper and hypothyroidism,
CC   multiple sclerosis, lung diseases including asthma, Crohn's disease,
CC   scleroderma, autoimmune diseases, developmental disorders and neural tube
CC   defects. The present sequence is the coding sequence for MOL10a.
CC   MOL10a protein is a cyclic nucleotide-gated olfactory channel-like
CC   protein.
XX
SQ   Sequence 1835 BP; 420 A; 521 C; 521 G; 373 T; 0 other;
```

alignment_scores:

```
            Quality: 2957.50              Length:    578
              Ratio:  5.143                Gaps:      3
 Percent Similarity:  99.481   Percent Identity: 99.481
```

alignment_block:

US-09-927-267-1 x ABA04602

Align seg 1/1 to: ABA04602 from: 1 to: 1835

```
   1 MetSerGlnAspThrLysValLysThrThrGluSerSerProProAlaPr 17
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  54 ATGAGCCAGGACACCAAAGTGAAGACAACAGAGTCCAGTCCCCCAGCCCC 103

  17 oSerLysAla...ArgLysLeuLeuProValLeuAspProSerGlyAspT 33
     ||||||||||   ||||||||||||||||||||||||||||||||||||
 104 ATCCAAGGCCAGGAGGAAGTTGCTGCCTGTCCTGGACCCATCTGGGGATT 153

  33 yrTyrTyrTrpTrpLeuAsnThrMetValPheProValMetTyrAsnLeu 49
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 154 ACTACTACTGGTGGCTGAACACAATGGTCTTCCCAGTCATGTATAACCTC 203

  50 IleIleLeuValCysArgAlaCysPheProAspLeuGlnHisGlyTyrLe 66
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 204 ATCATCCTCGTGTGCAGGGCCTGCTTCCCCGACTTGCAGCACGGTTATCT 253

  66 uValAlaTrpLeuValLeuAspTyrThrSerAspLeuLeuTyrLeuLeuA 83
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 254 GGTGGCCTGGTTGGTGCTGGACTACACGAGTGACCTGCTATACCTACTAG 303

  83 spMetValValArgPheHisThr...GlyPheLeuGluGlnGlyIleLeu 98
     ||||||||||||||||||||||   ||||||||||||||||||||||||
 304 ACATGGTGGTGCGCTTCCACACAGGTGGATTCTTGGAACAGGGCATCCTG 353
```

THIS PAGE BLANK (USPTO)

Query Match 99.9%; Score 1726.4; DB 6; Length 2013;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1727; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1 atgagccaggacaccaaagtgaagacaacagagtccagtcccccagccccatccaaggcc 60
Db     124 ATGAGCCAGGACACCAAAGTGAAGACAACAGAGTCCAGTCCCCCAGCCCCATCCAAGGCC 183

Qy      61 aggaagttgctgcctgtcctggacccatctggggattactactactggtggctgaacaca 120
Db     184 AGGAAGTTGCTGCCTGTCCTGGACCCATCTGGGGATTACTACTACTGGTGGCTGAACACA 243

Qy     121 atggtcttcccagtcatgtataacctcatcatcctcgtgtgcagagcctgcttccccgac 180
Db     244 ATGGTCTTCCCAGTCATGTATAACCTCATCATCCTCGTGTGCAGAGCCTGCTTCCCCGAC 303

Qy     181 ttgcagcacggttatctggtggcctggttggtgctggactacacgagtgacctgctatac 240
Db     304 TTGCAGCACGGTTATCTGGTGGCCTGGTTGGTGCTGGACTACACGAGTGACCTGCTATAC 363

Qy     241 ctactagacatggtggtgcgcttccacacaggattcttggaacagggcatcctggtggtg 300
Db     364 CTACTAGACATGGTGGTGCGCTTCCACACAGGATTCTTGGAACAGGGCATCCTGGTGGTG 423

Qy     301 gacaagggtaggatctcgagtcgctacgttcgcacctggagtttcttcttggacctggct 360
Db     424 GACAAGGGTAGGATCTCGAGTCGCTACGTTCGCACCTGGAGTTTCTTCTTGGACCTGGCT 483

Qy     361 tccctgatgcccacagatgtggtctacgtgcggctgggcccgcacacacccaccctgagg 420
Db     484 TCCCTGATGCCCACAGATGTGGTCTACGTGCGGCTGGGCCCGCACACACCCACCCTGAGG 543

Qy     421 ctgaaccgctttctccgcgcgcccccgcctcttcgaggccttcgaccgcacagagacccgc 480
Db     544 CTGAACCGCTTTCTCCGCGCGCCCCCGCCTCTTCGAGGCCTTCGACCGCACAGAGACCCGC 603

Qy     481 acagcttacccaaatgcctttcgcattgccaagctgatgctttacatttttgtcgtcatc 540
Db     604 ACAGCTTACCCAAATGCCTTTCGCATTGCCAAGCTGATGCTTTACATTTTTGTCGTCATC 663

Qy     541 cattggaacagctgcctatactttgccctatcccggtacctgggcttcgggcgtgacgca 600
Db     664 CATTGGAACAGCTGCCTATACTTTGCCCTATCCCGGTACCTGGGCTTCGGGCGTGACGCA 723

Qy     601 tgggtgtacccggaccccgcgcagcctggctttgagcgcctgcggcgccagtacctctat 660
Db     724 TGGGTGTACCCGGACCCCGCGCAGCCTGGCTTTGAGCGCCTGCGGCGCCAGTACCTCTAT 783

Qy     661 agcttttacttctccacgctgatactgactacagtgggcgatacaccgccgccagccagg 720
Db     784 AGCTTTTACTTCCCCACGCTGATACTGACTACAGTGGGCGATACACCGCCGCCAGCCAGG 843

Qy     721 gaagaagagtacctcttcatggtgggcgacttcctgctggccgtcatgggtttcgccacc 780
Db     844 GAAGAAGAGTACCTCTTCATGGTGGGCGACTTCCTGCTGGCCGTCATGGGTTTCGCCACC 903

Qy     781 atcatgggtagcatgagctctgtcatctacaacatgaacactgcagatgcggctttctac 840
Db     904 ATCATGGGTAGCATGAGCTCTGTCATCTACAACATGAACACTGCAGATGCGGCTTTCTAC 963

Qy     841 ccagatcatgcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggag 900
Db     964 CCAGATCATGCACTGGTGAAGAAGTACATGAAGCTGCAGCACGTCAACCGCAAGCTGGAG 1023

Qy     901 cggcgagttattgactggtatcagcacctgcagatcaacaagaagatgaccaacgaggta 960
Db    1024 CGGCGAGTTATTGACTGGTATCAGCACCTGCAGATCAACAAGAAGATGACCAACGAGGTA 1083

Qy     961 gccatcttacagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctg 1020
Db    1084 GCCATCTTACAGCACTTGCCTGAGCGGCTGCGGGCAGAAGTGGCTGTGTCTGTGCACCTG 1143

Qy    1021 tccactctgagccgggtgcagatctttcagaactgtgaggccagcctgctggaggagctg 1080
Db    1144 TCCACTCTGAGCCGGGTGCAGATCTTTCAGAACTGTGAGGCCAGCCTGCTGGAGGAGCTG 1203

Qy    1081 gtgctgaagctgcagccccagacctactcaccaggtgaatatgtatgccgcaaaggagac 1140
Db    1204 GTGCTGAAGCTGCAGCCCCAGACCTACTCACCAGGTGAATATGTATGCCGCAAAGGAGAC 1263

Qy    1141 attggccaagagatgtacatcatccgagagggtcaactggccgtggtggcagatgatggt 1200
Db    1264 ATTGGCCAAGAGATGTACATCATCCGAGAGGGTCAACTGGCCGTGGTGGCAGATGATGGT 1323

Qy    1201 atcacacagtatgctgtgctcggtgcagggctctactttggggagatcagcatcatcaac 1260
Db    1324 ATCACACAGTATGCTGTGCTCGGTGCAGGGCTCTACTTTGGGGAGATCAGCATCATCAAC 1383

Qy    1261 atcaaagggaacatgtctgggaaccgccgcacagccaacatcaagagcctaggttattca 1320
Db    1384 ATCAAAGGGAACATGTCTGGGAACCGCCGCACAGCCAACATCAAGAGCCTAGGTTATTCA 1443

Qy    1321 gacctattctgcctgagcaaggaggacctgcgggaggtgctgagcgagtatccacaagca 1380
Db    1444 GACCTATTCTGCCTGAGCAAGGAGGACCTGCGGGAGGTGCTGAGCGAGTATCCACAAGCA 1503

Qy    1381 cagaccatcatggaggagaaaggacgtgagatcctgctgaaaatgaacaagttggacgtg 1440
Db    1504 CAGACCATCATGGAGGAGAAAGGACGTGAGATCCTGCTGAAAATGAACAAGTTGGACGTG 1563

Qy    1441 aatgctgaggcagctgagatcgccctgcaggaggccacagagtcccggctacgaggccta 1500
Db    1564 AATGCTGAGGCAGCTGAGATCGCCCTGCAGGAGGCCACAGAGTCCCGGCTACGAGGCCTA 1623

Qy    1501 gaccagcagctggatgatctacagaccaagtttgctcgcctcctggctgagctggagtcc 1560
Db    1624 GACCAGCAGCTGGATGATCTACAGACCAAGTTTGCTCGCCTCCTGGCTGAGCTGGAGTCC 1683

Qy    1561 agcgcacttaagattgcttaccgcattgaacggctggagtggcagactcgagagtggcca 1620
Db    1684 AGCGCACTTAAGATTGCTTACCGCATTGAACGGCTGGAGTGGCAGACTCGAGAGTGGCCA 1743

Qy    1621 atgcccgaggacctggctgaggctgatgacgagggtgagcctgaggagggaacttccaaa 1680
Db    1744 ATGCCCGAGGACCTGGCTGAGGCTGATGACGAGGGTGAGCCTGAGGAGGGAACTTCCAAA 1803

Qy    1681 gatgaagagggcagggccagccaggagggaccccccaggtccagagtga 1728
Db    1804 GATGAAGAGGGCAGGGCCAGCCAGGAGGGACCCCCAGGTCCAGAGTGA 1851
```

RESULT 2
AX350404
LOCUS AX350404 2551 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 29 from Patent WO0181578.
ACCESSION AX350404
VERSION AX350404.1 GI:18616046
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Vernet,C.A., Fernandes,E.R., Gerlach,V., Shimkets,R.A.,
Malyankar,U.M., Boldog,F.L., Zerhusen,B.D., Spytek,K.A.,
Majumder,K., Tchernev,V.T., Padigaru,M., Patturajan,M.,
Burgess,C.E., Gangolli,E.A., Smithson,G., Rastelli,L.,
Macdougall,J.R., Taupier,R.J., Grosse,W.M. and Alsobrook,J.P.
TITLE Novel proteins and nucleic acids encoding same
JOURNAL Patent: WO 0181578-A 29 01-NOV-2001;
Curagen Corporation (US)
FEATURES Location/Qualifiers
source 1. .2551
/organism="Homo sapiens"

THIS PAGE BLANK (USPTO)

/db_xref="taxon:9606"
BASE COUNT 646 a 668 c 707 g 530 t
ORIGIN

Query Match 99.9%; Score 1726.4; DB 6; Length 2551;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1727; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 atgagccaggacaccaaagtgaagacaacagagtccagtcccccagccccatccaaggcc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    779 ATGAGCCAGGACACCAAAGTGAAGACAACAGAGTCCAGTCCCCCAGCCCCATCCAAGGCC 838

Qy     61 aggaagttgctgcctgtcctggacccatctggggattactactactggtggctgaacaca 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    839 AGGAAGTTGCTGCCTGTCCTGGACCCATCTGGGGATTACTACTACTGGTGGCTGAACACA 898

Qy    121 atggtcttcccagtcatgtataacctcatcatcctcgtgtgcagagcctgcttccccgac 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    899 ATGGTCTTCCCAGTCATGTATAACCTCATCATCCTCGTGTGCAGAGCCTGCTTCCCCGAC 958

Qy    181 ttgcagcacggttatctggtggcctggttggtgctggactacacgagtgacctgctatac 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    959 TTGCAGCACGGTTATCTGGTGGCCTGGTTGGTGCTGGACTACACGAGTGACCTGCTATAC 1018

Qy    241 ctactagacatggtggtgcgcttccacacaggattcttggaacagggcatcctggtggtg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1019 CTACTAGACATGGTGGTGCGCTTCCACACAGGATTCTTGGAACAGGGCATCCTGGTGGTG 1078

Qy    301 gacaagggtaggatctcgagtcgctacgttcgcacctggagtttcttcttggacctggct 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1079 GACAAGGGTAGGATCTCGAGTCGCTACGTTCGCACCTGGAGTTTCTTCTTGGACCTGGCT 1138

Qy    361 tccctgatgcccacagatgtggtctacgtgcggctgggcccgcacacacccaccctgagg 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1139 TCCCTGATGCCCACAGATGTGGTCTACGTGCGGCTGGGCCCGCACACACCCACCCTGAGG 1198

Qy    421 ctgaaccgctttctccgcgcgccccgcctcttcgaggccttcgaccgcacagagacccgc 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1199 CTGAACCGCTTTCTCCGCGCGCCCCCGCCTCTTCGAGGCCTTCGACCGCGCAGAGACCCGC 1258

Qy    481 acagcttacccaaatgcctttcgcattgccaagctgatgctttacatttttgtcgtcatc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1259 ACAGCTTACCCAAATGCCTTTCGCATTGCCAAGCTGATGCTTTACATTTTTGTCGTCATC 1318

Qy    541 cattggaacagctgcctatactttgccctatcccggtacctgggcttcgggcgtgacgca 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1319 CATTGGAACAGCTGCCTATACTTTGCCCTATCCCGGTACCTGGGCTTCGGGCGTGACGCA 1378

Qy    601 tgggtgtacccggaccccgcgcagcctggctttgagcgcctgcggcgccagtacctctat 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1379 TGGGTGTACCCGGACCCCGCGCAGCCTGGCTTTGAGCGCCTGCGGCGCCAGTACCTCTAT 1438

Qy    661 agcttttacttctccacgctgatactgactacagtgggcgatacaccgccgccagccagg 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1439 AGCTTTTACTTCTCCACGCTGATACTGACTACAGTGGGCGATACACCGCCGCCAGCCAGG 1498

Qy    721 gaagaagagtacctcttcatggtgggcgacttcctgctggccgtcatgggtttcgccacc 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1499 GAAGAAGAGTACCTCTTCATGGTGGGCGACTTCCTGCTGGCCGTCATGGGTTTCGCCACC 1558

Qy    781 atcatgggtagcatgagctctgtcatctacaacatgaacactgcagatgcggctttctac 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1559 ATCATGGGTAGCATGAGCTCTGTCATCTACAACATGAACACTGCAGATGCGGCTTTCTAC 1618

Qy    841 ccagatcatgcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggag 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1619 CCAGATCATGCACTGGTGAAGAAGTACATGAAGCTGCAGCACGTCAACCGCAAGCTGGAG 1678

Qy    901 cggcgagttattgactggtatcagcacctgcagatcaacaagaagatgaccaacgaggta 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1679 CGGCGAGTTATTGACTGGTATCAGCACCTGCAGATCAACAAGAAGATGACCAACGAGGTA 1738
```

```
Qy    961 gccatcttacagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctg 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1739 GCCATCTTACAGCACTTGCCTGAGCGGCTGCGGGCAGAAGTGGCTGTGTCTGTGCACCTG 1798

Qy   1021 tccactctgagccgggtgcagatctttcagaactgtgaggccagcctgctggaggagctg 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1799 TCCACTCTGAGCCGGGTGCAGATCTTTCAGAACTGTGAGGCCAGCCTGCTGGAGGAGCTG 1858

Qy   1081 gtgctgaagctgcagccccagacctactcaccaggtgaatatgtatgccgcaaaggagac 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1859 GTGCTGAAGCTGCAGCCCCAGACCTACTCACCAGGTGAATATGTATGCCGCAAAGGAGAC 1918

Qy   1141 attggccaagagatgtacatcatccgagagggtcaactggccgtggtggcagatgatggt 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1919 ATTGGCCAAGAGATGTACATCATCCGAGAGGGTCAACTGGCCGTGGTGGCAGATGATGGT 1978

Qy   1201 atcacacagtatgctgtgctcggtgcagggctctactttggggagatcagcatcatcaac 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1979 ATCACACAGTATGCTGTGCTCGGTGCAGGGCTCTACTTTGGGGAGATCAGCATCATCAAC 2038

Qy   1261 atcaaagggaacatgtctgggaaccgccgcacagccaacatcaagagcctaggttattca 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2039 ATCAAAGGGAACATGTCTGGGAACCGCCGCACAGCCAACATCAAGAGCCTAGGTTATTCA 2098

Qy   1321 gacctattctgcctgagcaaggaggacctgcgggaggtgctgagcgagtatccacaagca 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2099 GACCTATTCTGCCTGAGCAAGGAGGACCTGCGGGAGGTGCTGAGCGAGTATCCACAAGCA 2158

Qy   1381 cagaccatcatggaggagaaaggacgtgagatcctgctgaaaatgaacaagttggacgtg 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2159 CAGACCATCATGGAGGAGAAAGGACGTGAGATCCTGCTGAAAATGAACAAGTTGGACGTG 2218

Qy   1441 aatgctgaggcagctgagatcgccctgcaggaggccacagagtcccggctacgaggccta 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2219 AATGCTGAGGCAGCTGAGATCGCCCTGCAGGAGGCCACAGAGTCCCGGCTACGAGGCCTA 2278

Qy   1501 gaccagcagctggatgatctacagaccaagtttgctcgcctcctggctgagctggagtcc 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2279 GACCAGCAGCTGGATGATCTACAGACCAAGTTTGCTCGCCTCCTGGCTGAGCTGGAGTCC 2338

Qy   1561 agcgcacttaagattgcttaccgcattgaacggctggagtggcagactcgagagtggcca 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2339 AGCGCACTTAAGATTGCTTACCGCATTGAACGGCTGGAGTGGCAGACTCGAGAGTGGCCA 2398

Qy   1621 atgcccgaggacctggctgaggctgatgacgagggtgagcctgaggagggaacttccaaa 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2399 ATGCCCGAGGACCTGGCTGAGGCTGATGACGAGGGTGAGCCTGAGGAGGGAACTTCCAAA 2458

Qy   1681 gatgaagagggcagggccagccaggagggacccccaggtccagagtga 1728
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   2459 GATGAAGAGGGCAGGGCCAGCCAGGAGGGACCCCCAGGTCCAGAGTGA 2506
```

RESULT 3
AX350402
LOCUS AX350402 1835 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 27 from Patent WO0181578.
ACCESSION AX350402
VERSION AX350402.1 GI:18616045
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Vernet,C.A., Fernandes,E.R., Gerlach,V., Shimkets,R.A.,
Malyankar,U.M., Boldog,F.L., Zerhusen,B.D., Spytek,K.A.,
Majumder,K., Tchernev,V.T., Padigaru,M., Patturajan,M.,
Burgess,C.E., Gangolli,E.A., Smithson,G., Rastelli,L.,
Macdougall,J.R., Taupier,R.J., Grosse,W.M. and Alsobrook,J.P.
TITLE Novel proteins and nucleic acids encoding same
JOURNAL Patent: WO 0181578-A 27 01-NOV-2001;
Curagen Corporation (US)

THIS PAGE BLANK (USPTO)

```
OM of: US-09-927-267-1 to: GenEmbl:*   out_format : pfs

Date: Jun 20, 2002  6:29 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09927267/runat_19062002_090142_25489/app_query.fasta_1.640
 -DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500
 -MINLEN=0 -MAXLEN=2000000000 -USER=US09927267_@CGN1_1_8051
 -NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-927-267-1
Query length: 575
Database: GenEmbl:*
Database sequences: 1797656
Database length: 1873333701
Search time (sec): 2604.520000


score_list:
Sequence           Strd Orig    ZScore    EScore Len  ! Documentation  ..
gb_pat:AX355967      + 2984.00 4706.27  8.9e-254  2013  ! AX355967 Sequence 2 from Paten
gb_pat:AX350404      + 2984.00 4703.76  1.2e-253  2551  ! AX350404 Sequence 29 from Pate
gb_pat:AX350402      + 2957.50 4665.26  1.7e-251  1835  ! AX350402 Sequence 27 from Pate
gb_ro:RNU12425       + 2811.00 4432.56  1.6e-238  1938  ! U12425 Rattus norvegicus olfac
gb_ro:RNU12623       + 2811.00 4430.51  2.0e-238  2351  ! U12623 Rattus norvegicus Sprag
gb_htg:AC021935      + 2266.00 3521.97  8.2e-188  162965  ! AC021935 Homo sapiens chromo
gb_htg:AC022762      + 2266.00 3520.53  9.9e-188  186609  ! AC022762 Homo sapiens chromo
gb_htg:AC036216      + 2257.00 3507.61  5.2e-187  164434  ! AC036216 Homo sapiens chromo
gb_ro:MMU243933      + 1580.50 2479.02  1.0e-129  2781  ! AJ243933 Mus musculus mRNA for
gb_om:BTCNGCHAN      + 1579.00 2477.20  1.3e-129  2641  ! X89600 B.taurus mRNA for alpha
gb_om:BTCNG3         + 1579.00 2472.79  2.3e-129  3998  ! X76485 B.taurus CNG3 mRNA for
gb_ro:RNO272429      + 1578.00 2478.06  1.1e-129  2098  ! AJ272429 Rattus norvegicus mRN
gb_ro:RNO272428      + 1578.00 2477.49  1.2e-129  2212  ! AJ272428 Rattus norvegicus mRN
gb_ro:AB002801       + 1578.00 2477.49  1.2e-129  2213  ! AB002801 Rattus norvegicus mRN
gb_om:BTCAMPGC       + 1568.00 2457.84  1.5e-128  3166  ! X55010 B.taurus mRNA for cAMP-
gb_ov:GGCONECNG      + 1562.50 2450.83  3.8e-128  2697  ! X89598 G.gallus mRNA for alpha
gb_om:OCRACNG        + 1562.00 2450.15  4.1e-128  2669  ! X59668 O.cuniculus rACNG mRNA
gb_ro:AF126808       + 1558.00 2443.24  1.0e-127  2615  ! AF126808 Rattus norvegicus olf
gb_ro:RNOLFCH        + 1558.00 2442.47  1.1e-127  3027  ! X55519 Rat mRNA for a cyclic n
gb_pr:AF065314       + 1549.50 2427.51  7.5e-127  3486  ! AF065314 Homo sapiens cone pho
gb_om:BTCGMPCH       + 1544.50 2422.37  1.5e-126  2682  ! X51604 B.taurus RNA for cGMP-g
gb_ro:MMU19717       + 1543.00 2421.64  1.6e-126  2297  ! U19717 Mus musculus cGMP-gated
gb_om:CFACGMP        + 1541.50 2417.56  2.7e-126  2697  ! X99914 C.familiaris gene encod
gb_om:CFU83905       + 1541.50 2417.48  2.7e-126  2717  ! U83905 Canis familiaris rod ph
gb_om:SSU85404       + 1538.50 2415.66  3.4e-126  2061  ! U85404 Sus scrofa cGMP-gated c
gb_pr:S42457         + 1538.50 2412.19  5.4e-126  2857  ! S42457 CNCG-rod photoreceptor
gb_ro:RNU93851       + 1537.00 2412.84  4.9e-126  2148  ! U93851 Rattus norvegicus rod-t
gb_pat:AX179746      + 1535.00 2410.46  6.7e-126  1995  ! AX179746 Sequence 34 from Pate
gb_ro:MUSCNCG        + 1534.00 2406.59  1.1e-125  2474  ! M84742 Mus musculus cGMP-gated
gb_ro:RNU48803       + 1532.00 2405.40  1.3e-125  2052  ! U48803 Rattus norvegicus rod p
gb_ov:GGRODCNG       + 1529.00 2398.09  3.3e-125  2612  ! X89599 G.gallus mRNA for alpha
gb_pr:HUMCGMP        + 1526.50 2394.59  5.1e-125  2500  ! M84741 Human cGMP-gated cation
gb_ov:ICTCNC         + 1436.50 2254.10  3.4e-117  2049  ! M83111 Ictalurus punctatus cyc
gb_htg:AC091450      + 1399.00 2148.51  2.6e-111  158588  ! AC091450 Mus musculus chromo
gb_ro:MMU49391       + 1397.00 2175.10  8.6e-113  9618  ! U49391 Mus musculus cyclic nuc
gb_ov:AF393839       + 1380.50 2159.05  6.7e-112  3718  ! AF393839 Oncorhynchus mykiss c
gb_pr:HSAF002992     + 1355.00 2083.27  1.1e-107  104037  ! AF002992 Homo sapiens cosmid
gb_ro:MMU238241      + 1349.50 2120.09  1.0e-109  1428  ! AJ238241 Mus musculus CNG3 gen
gb_pr:AC092675       + 1344.00 2061.71  1.8e-106  153551  ! AC092675 Homo sapiens chromo
gb_htg:AL671880      + 1336.00 2046.23  1.3e-105  200039  ! AL671880 Mus musculus chromo
gb_htg:AC099698      - 1336.00 2045.27  1.5e-105  218872  ! AC099698 Mus musculus chr
em_htg_hum:AC013535  + 1323.00 2024.51  2.1e-104  222364  ! AC013535 Homo sapiens chr
gb_htg:AC107068      - 1323.00 2024.16  2.2e-104  229687  ! AC107068 Homo sapiens chr
gb_in:DMCNGCHAN      + 1241.50 1941.39  9.0e-100  2913  ! X89601 D.melanogaster mRNA
gb_htg:AC007704      - 1210.50 1847.92  1.4e-94   190000  ! AC007704 Homo sapiens chr

seq_name: gb_pat:AX355967

seq_documentation_block:
LOCUS       AX355967               2013 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 2 from Patent WO0202633.
ACCESSION   AX355967
VERSION     AX355967.1  GI:18620561
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Raumann,B.E., Sanjanwala,M.S., Tribouley,C.M. and Walia,N.K.
  TITLE     Transporters and ion channels
  JOURNAL   Patent: WO 0202633-A 2 10-JAN-2002;
            Incyte Genomics, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .2013
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /note="Incyte ID No: 6703242CB1"
BASE COUNT      472 a    576 c    557 g    408 t
ORIGIN


alignment_scores:
            Quality: 2984.00              Length:    575
              Ratio:  5.199                 Gaps:      0
Percent Similarity:  99.826   Percent Identity: 99.826

alignment_block:
 US-09-927-267-1 x AX355967   ..

 Align seg 1/1  to: AX355967  from: 1  to: 2013

       1 MetSerGlnAspThrLysValLysThrThrGluSerSerProProAlaPr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     124 ATGAGCCAGGACACCAAAGTGAAGACAACAGAGTCCAGTCCCCCAGCCCC 173

      17 oSerLysAlaArgLysLeuLeuProValLeuAspProSerGlyAspTyrT 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     174 ATCCAAGGCCAGGAAGTTGCTGCCTGTCCTGGACCCATCTGGGGATTACT 223

      34 yrTyrTrpTrpLeuAsnThrMetValPheProValMetTyrAsnLeuIle 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     224 ACTACTGGTGGCTGAACACAATGGTCTTCCCAGTCATGTATAACCTCATC 273

      51 IleLeuValCysArgAlaCysPheProAspLeuGlnHisGlyTyrLeuVa 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     274 ATCCTCGTGTGCAGAGCCTGCTTCCCCGACTTGCAGCACGGTTATCTGGT 323

      67 lAlaTrpLeuValLeuAspTyrThrSerAspLeuLeuTyrLeuLeuAspM 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     324 GGCCTGGTTGGTGCTGGACTACACGAGTGACCTGCTATACCTACTAGACA 373

      84 etValValArgPheHisThrGlyPheLeuGluGlnGlyIleLeuValVal 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     374 TGGTGGTGCGCTTCCACACAGGATTCTTGGAACAGGGCATCCTGGTGGTG 423

     101 AspLysGlyArgIleSerSerArgTyrValArgThrTrpSerPhePheLe 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     424 GACAAGGGTAGGATCTCGAGTCGCTACGTTCGCACCTGGAGTTTCTTCTT 473

     117 uAspLeuAlaSerLeuMetProThrAspValValTyrValArgLeuGlyP 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     474 GGACCTGGCTTCCCTGATGCCCACAGATGTGGTCTACGTGCGGCTGGGCC 523
```

THIS PAGE BLANK (USPTO)

```
 134 roHisThrProThrLeuArgLeuAsnArgPheLeuArgAlaProArgLeu 150
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 524 CGCACACACCCACCCTGAGGCTGAACCGCTTTCTCCGCGCGCCCCGCCTC 573

 151 PheGluAlaPheAspArgThrGluThrArgThrAlaTyrProAsnAlaPh 167
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 574 TTCGAGGCCTTCGACCGCACAGAGACCCGCACAGCTTACCCAAATGCCTT 623

 167 eArgIleAlaLysLeuMetLeuTyrIlePheValValIleHisTrpAsnS 184
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 624 TCGCATTGCCAAGCTGATGCTTTACATTTTTGTCGTCATCCATTGGAACA 673

 184 erCysLeuTyrPheAlaLeuSerArgTyrLeuGlyPheGlyArgAspAla 200
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 674 GCTGCCTATACTTTGCCCTATCCCGGTACCTGGGCTTCGGGCGTGACGCA 723

 201 TrpValTyrProAspProAlaGlnProGlyPheGluArgLeuArgArgGl 217
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 724 TGGGTGTACCCGGACCCCGCGCAGCCTGGCTTTGAGCGCCTGCGGCGCCA 773

 217 nTyrLeuTyrSerPheTyrPheThrLeuIleLeuThrThrValGlyA 234
     ||||||||||||||||||||||||||||||||||||||||||||||||
 774 GTACCTCTATAGCTTTTACTTGCCCACGCTGATACTGACTACAGTGGGCG 823

 234 spThrProProProAlaArgGluGluGluTyrLeuPheMetValGlyAsp 250
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 824 ATACACCGCCGCCAGCCAGGGAAGAAGAGTACCTCTTCATGGTGGGCGAC 873

 251 PheLeuLeuAlaValMetGlyPheAlaThrIleMetGlySerMetSerSe 267
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 874 TTCCTGCTGGCCGTCATGGGTTTCGCCACCATCATGGGTAGCATGAGCTC 923

 267 rValIleTyrAsnMetAsnThrAlaAspAlaAlaPheTyrProAspHisA 284
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 924 TGTCATCTACAACATGAACACTGCAGATGCGGCTTTCTACCCAGATCATG 973

 284 laLeuValLysLysTyrMetLysLeuGlnHisValAsnArgLysLeuGlu 300
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 974 CACTGGTGAAGAAGTACATGAAGCTGCAGCACGTCAACCGCAAGCTGGAG 1023

 301 ArgArgValIleAspTrpTyrGlnHisLeuGlnIleAsnLysLysMetTh 317
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1024 CGGCGAGTTATTGACTGGTATCAGCACCTGCAGATCAACAAGAAGATGAC 1073

 317 rAsnGluValAlaIleLeuGlnHisLeuProGluArgLeuArgAlaGluV 334
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1074 CAACGAGGTAGCCATCTTACAGCACTTGCCTGAGCGGCTGCGGGCAGAAG 1123

 334 alAlaValSerValHisLeuSerThrLeuSerArgValGlnIlePheGln 350
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1124 TGGCTGTGTCTGTGCACCTGTCCACTCTGAGCCGGGTGCAGATCTTTCAG 1173

 351 AsnCysGluAlaSerLeuLeuGluGluLeuValLeuLysLeuGlnProGl 367
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1174 AACTGTGAGGCCAGCCTGCTGGAGGAGCTGGTGCTGAAGCTGCAGCCCCA 1223

 367 nThrTyrSerProGlyGluTyrValCysArgLysGlyAspIleGlyGlnG 384
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1224 GACCTACTCACCAGGTGAATATGTATGCCGCAAAGGAGACATTGGCCAAG 1273

 384 luMetTyrIleIleArgGluGlyGlnLeuAlaValValAlaAspAspGly 400
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1274 AGATGTACATCATCCGAGAGGGTCAACTGGCCGTGGTGGCAGATGATGGT 1323

 401 IleThrGlnTyrAlaValLeuGlyAlaGlyLeuTyrPheGlyGluIleSe 417
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1324 ATCACACAGTATGCTGTGCTCGGTGCAGGGCTCTACTTTGGGGAGATCAG 1373

 417 rIleIleAsnIleLysGlyAsnMetSerGlyAsnArgArgThrAlaAsnI 434
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1374 CATCATCAACATCAAAGGGAACATGTCTGGGAACCGCCGCACAGCCAACA 1423

 434 leLysSerLeuGlyTyrSerAspLeuPheCysLeuSerLysGluAspLeu 450
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1424 TCAAGAGCCTAGGTTATTCAGACCTATTCTGCCTGAGCAAGGAGGACCTG 1473

 451 ArgGluValLeuSerGluTyrProGlnAlaGlnThrIleMetGluGluLy 467
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1474 CGGGAGGTGCTGAGCGAGTATCCACAAGCACAGACCATCATGGAGGAGAA 1523

 467 sGlyArgGluIleLeuLeuLysMetAsnLysLeuAspValAsnAlaGluA 484
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1524 AGGACGTGAGATCCTGCTGAAAATGAACAAGTTGGACGTGAATGCTGAGG 1573

 484 laAlaGluIleAlaLeuGlnGluAlaThrGluSerArgLeuArgGlyLeu 500
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1574 CAGCTGAGATCGCCCTGCAGGAGGCCACAGAGTCCCGGCTACGAGGCCTA 1623

 501 AspGlnGlnLeuAspAspLeuGlnThrLysPheAlaArgLeuLeuAlaGl 517
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1624 GACCAGCAGCTGGATGATCTACAGACCAAGTTTGCTCGCCTCCTGGCTGA 1673

 517 uLeuGluSerSerAlaLeuLysIleAlaTyrArgIleGluArgLeuGluT 534
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1674 GCTGGAGTCCAGCGCACTTAAGATTGCTTACCGCATTGAACGGCTGGAGT 1723

 534 rpGlnThrArgGluTrpProMetProGluAspLeuAlaGluAlaAspAsp 550
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1724 GGCAGACTCGAGAGTGGCCAATGCCCGAGGACCTGGCTGAGGCTGATGAC 1773

 551 GluGlyGluProGluGluGlyThrSerLysAspGluGluGlyArgAlaSe 567
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1774 GAGGGTGAGCCTGAGGAGGGAACTTCCAAAGATGAAGAGGGCAGGGCCAG 1823

 567 rGlnGluGlyProProGlyProGlu 575
     ||||||||||||||||||||||||
1824 CCAGGAGGGACCCCCAGGTCCAGAG 1848
```

```
seq_name: gb_pat:AX350404

seq_documentation_block:
LOCUS       AX350404               2551 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 29 from Patent WO0181578.
ACCESSION   AX350404
VERSION     AX350404.1  GI:18616046
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Vernet,C.A., Fernandes,E.R., Gerlach,V., Shimkets,R.A.,
            Malyankar,U.M., Boldog,F.L., Zerhusen,B.D., Spytek,K.A.,
            Majumder,K., Tchernev,V.T., Padigaru,M., Patturajan,M.,
            Burgess,C.E., Gangolli,E.A., Smithson,G., Rastelli,L.,
            Macdougall,J.R., Taupier,R.J., Grosse,W.M. and Alsobrook,J.P.
  TITLE     Novel proteins and nucleic acids encoding same
  JOURNAL   Patent: WO 0181578-A 29 01-NOV-2001;
            Curagen Corporation (US)
FEATURES             Location/Qualifiers
     source          1..2551
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      646 a    608 c    707 g    530 t
ORIGIN


alignment_scores:
            Quality: 2984.00              Length:    575
              Ratio:  5.190                Gaps:      0
 Percent Similarity: 100.000  Percent Identity: 99.826

alignment_block:
 US-09-927-267-1 x AX350404   ..
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2002, 14:30:23 ; Search time 7623.77 Seconds
(without alignments)
4743.199 Million cell updates/sec

Title: US-09-927-267-3
Perfect score: 1728
Sequence: 1 atgagccaggacaccaaagt.....gacccccaggtccagagtga 1728

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1726.4	99.9	2013	6	AX355967	AX355967 Sequence
2	1726.4	99.9	2551	6	AX350404	AX350404 Sequence
3	1687.4	97.7	1835	6	AX350402	AX350402 Sequence
4	1420.8	82.2	1938	10	RNU12425	U12425 Rattus norv
5	1420.8	82.2	2351	10	RNU12623	U12623 Rattus norv
6	652.6	37.8	162965	2	AC021935	AC021935 Homo sapi
7	652.6	37.8	164434	2	AC036216	AC036216 Homo sapi
8	652.6	37.8	186609	2	AC022762	AC022762 Homo sapi
9	617.4	35.7	2641	4	BTCNGCHAN	X89600 B.taurus mR
10	617.4	35.7	3998	4	BTCNG3	X76485 B.taurus CN
11	588.2	34.0	3486	9	AF065314	AF065314 Homo sapi
12	584.8	33.8	2781	10	MMU243933	AJ243933 Mus muscu
13	583.2	33.8	3166	4	BTCAMPGC	X55010 B.taurus mR
14	578.6	33.5	1995	6	AX179746	AX179746 Sequence
15	572	33.1	2669	4	OCRACNG	X59668 O.cuniculus
16	570	33.0	2098	10	RNO272429	AJ272429 Rattus no
17	570	33.0	2212	10	RNO272428	AJ272428 Rattus no
18	570	33.0	2213	10	AB002801	AB002801 Rattus no
19	530	30.7	2815	10	AF126808	AF126808 Rattus no
20	530	30.7	3027	10	RNOLFCH	X55519 Rat mRNA fo
21	515.2	29.8	153551	9	AC092675	AC092675 Homo sapi
22	511	29.6	1428	10	MMU238241	AJ238241 Mus muscu
23	507.4	29.4	2697	5	GGCONECNG	X89598 G.gallus mR
24	499.2	28.9	3718	5	AF393839	AF393839 Oncorhync
25	498.6	28.9	104037	9	HSAF002992	AF002992 Homo sapi
26	494.8	28.6	2049	5	ICTCNC	M83111 Ictalurus p
27	490	28.4	2052	10	RNU48803	U48803 Rattus norv
28	481.2	27.8	2148	10	RNU93851	U93851 Rattus norv
29	475.6	27.5	2682	4	BTCGMPCH	X51604 B.taurus RN
30	475.4	27.5	158588	2	AC091450	AC091450 Mus muscu
31	472.2	27.3	9618	10	MMU49391	U49391 Mus musculu
32	467.6	27.1	2697	4	CFACGMP	X99914 C.familiari
33	466	27.0	2297	10	MMU19717	U19717 Mus musculu
34	466	27.0	2717	4	CFU83905	U83905 Canis famil
35	464.4	26.9	2474	10	MUSCNCG	M84742 Mus musculu
36	461.2	26.7	2061	4	SSU85404	U85404 Sus scrofa
37	461.2	26.7	2857	9	S42457	S42457 CNCG=rod ph
38	459	26.6	2612	5	GGRODCNG	X89599 G.gallus mR
39	452.4	26.2	2500	9	HUMCGMP	M84741 Human cGMP-
40	423.2	24.5	200039	2	AL671880	AL671880 Mus muscu
c 41	423.2	24.5	218872	2	AC099698	AC099698 Mus muscu
42	416.8	24.1	222364	30	AC013535	AC013535 Homo sapi
c 43	416.8	24.1	229687	2	AC107068	AC107068 Homo sapi
44	410	23.7	2913	3	DMCNGCHAN	X89601 D.melanogas
c 45	386.4	22.4	190000	2	AC007704	AC007704 Homo sapi

ALIGNMENTS

RESULT 1
AX355967
LOCUS AX355967 2013 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 2 from Patent WO0202633.
ACCESSION AX355967
VERSION AX355967.1 GI:18620561
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Raumann,B.E., Sanjanwala,M.S., Tribouley,C.M. and Walia,N.K.
TITLE Transporters and ion channels
JOURNAL Patent: WO 0202633-A 2 10-JAN-2002;
Incyte Genomics, Inc. (US)
FEATURES Location/Qualifiers
source 1..2013
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Incyte ID No: 6703242CB1"
BASE COUNT 472 a 576 c 557 g 408 t
ORIGIN

THIS PAGE BLANK (USPTO)

Query Match 99.9%; Score 1726.4; DB 6; Length 2013;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1727; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 atgagccaggacaccaaagtgaagacaacagagtccagtcccccagccccatccaaggcc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    124 ATGAGCCAGGACACCAAAGTGAAGACAACAGAGTCCAGTCCCCCAGCCCCATCCAAGGCC 183

Qy     61 aggaagttgctgcctgtcctggacccatctggggattactactactggtggctgaacaca 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    184 AGGAAGTTGCTGCCTGTCCTGGACCCATCTGGGGATTACTACTACTGGTGGCTGAACACA 243

Qy    121 atggtcttcccagtcatgtataacctcatcatcctcgtgtgcagagcctgcttccccgac 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    244 ATGGTCTTCCCAGTCATGTATAACCTCATCATCCTCGTGTGCAGAGCCTGCTTCCCCGAC 303

Qy    181 ttgcagcacggttatctggtggcctggttggtgctggactacacgagtgacctgctatac 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    304 TTGCAGCACGGTTATCTGGTGGCCTGGTTGGTGCTGGACTACACGAGTGACCTGCTATAC 363

Qy    241 ctactagacatggtggtgcgcttccacacaggattcttggaacagggcatcctggtggtg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    364 CTACTAGACATGGTGGTGCGCTTCCACACAGGATTCTTGGAACAGGGCATCCTGGTGGTG 423

Qy    301 gacaagggtaggatctcgagtcgctacgttcgcacctggagtttcttcttggacctggct 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    424 GACAAGGGTAGGATCTCGAGTCGCTACGTTCGCACCTGGAGTTTCTTCTTGGACCTGGCT 483

Qy    361 tccctgatgcccacagatgtggtctacgtgcggctgggcccgcacacacccaccctgagg 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    484 TCCCTGATGCCCACAGATGTGGTCTACGTGCGGCTGGGCCCGCACACACCCACCCTGAGG 543

Qy    421 ctgaaccgctttctccgcgcgccccgcctcttcgaggccttcgaccgcacagagacccgc 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    544 CTGAACCGCTTTCTCCGCGCGCCCCGCCTCTTCGAGGCCTTCGACCGCACAGAGACCCGC 603

Qy    481 acagcttacccaaatgcctttcgcattgccaagctgatgctttacatttttgtcgtcatc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    604 ACAGCTTACCCAAATGCCTTTCGCATTGCCAAGCTGATGCTTTACATTTTTGTCGTCATC 663

Qy    541 cattggaacagctgcctatactttgccctatcccggtacctgggcttcgggcgtgacgca 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    664 CATTGGAACAGCTGCCTATACTTTGCCCTATCCCGGTACCTGGGCTTCGGGCGTGACGCA 723

Qy    601 tgggtgtacccggaccccgcgcagcctggctttgagcgcctgcggcgccagtacctctat 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    724 TGGGTGTACCCGGACCCCGCGCAGCCTGGCTTTGAGCGCCTGCGGCGCCAGTACCTCTAT 783

Qy    661 agcttttacttctccacgctgatactgactacagtgggcgatacaccgccgccagccagg 720
          ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    784 AGCTTTTACTTCCCCACGCTGATACTGACTACAGTGGGCGATACACCGCCGCCAGCCAGG 843

Qy    721 gaagaagagtacctcttcatggtgggcgacttcctgctggccgtcatgggtttcgccacc 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    844 GAAGAAGAGTACCTCTTCATGGTGGGCGACTTCCTGCTGGCCGTCATGGGTTTCGCCACC 903

Qy    781 atcatgggtagcatgagctctgtcatctacaacatgaacactgcagatgcggctttctac 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    904 ATCATGGGTAGCATGAGCTCTGTCATCTACAACATGAACACTGCAGATGCGGCTTTCTAC 963

Qy    841 ccagatcatgcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggag 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    964 CCAGATCATGCACTGGTGAAGAAGTACATGAAGCTGCAGCACGTCAACCGCAAGCTGGAG 1023

Qy    901 cggcgagttattgactggtatcagcacctgcagatcaacaagaagatgaccaacgaggta 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1024 CGGCGAGTTATTGACTGGTATCAGCACCTGCAGATCAACAAGAAGATGACCAACGAGGTA 1083

Qy    961 gccatcttacagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctg 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1084 GCCATCTTACAGCACTTGCCTGAGCGGCTGCGGGCAGAAGTGGCTGTGTCTGTGCACCTG 1143
```



```
Qy   1021 tccactctgagccgggtgcagatctttcagaactgtgaggccagcctgctggaggagctg 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1144 TCCACTCTGAGCCGGGTGCAGATCTTTCAGAACTGTGAGGCCAGCCTGCTGGAGGAGCTG 1203

Qy   1081 gtgctgaagctgcagccccagacctactcaccaggtgaatatgtatgccgcaaaggagac 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1204 GTGCTGAAGCTGCAGCCCCAGACCTACTCACCAGGTGAATATGTATGCCGCAAAGGAGAC 1263

Qy   1141 attggccaagagatgtacatcatccgagagggtcaactggccgtggtggcagatgatggt 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1264 ATTGGCCAAGAGATGTACATCATCCGAGAGGGTCAACTGGCCGTGGTGGCAGATGATGGT 1323

Qy   1201 atcacacagtatgctgtgctcggtgcagggctctactttggggagatcagcatcatcaac 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1324 ATCACACAGTATGCTGTGCTCGGTGCAGGGCTCTACTTTGGGGAGATCAGCATCATCAAC 1383

Qy   1261 atcaaagggaacatgtctgggaaccgccgcacagccaacatcaagagcctaggttattca 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1384 ATCAAAGGGAACATGTCTGGGAACCGCCGCACAGCCAACATCAAGAGCCTAGGTTATTCA 1443

Qy   1321 gacctattctgcctgagcaaggaggacctgcgggaggtgctgagcgagtatccacaagca 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1444 GACCTATTCTGCCTGAGCAAGGAGGACCTGCGGGAGGTGCTGAGCGAGTATCCACAAGCA 1503

Qy   1381 cagaccatcatggaggagaaaggacgtgagatcctgctgaaaatgaacaagttggacgtg 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1504 CAGACCATCATGGAGGAGAAAGGACGTGAGATCCTGCTGAAAATGAACAAGTTGGACGTG 1563

Qy   1441 aatgctgaggcagctgagatcgccctgcaggaggccacagagtcccggctacgaggccta 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1564 AATGCTGAGGCAGCTGAGATCGCCCTGCAGGAGGCCACAGAGTCCCGGCTACGAGGCCTA 1623

Qy   1501 gaccagcagctggatgatctacagaccaagtttgctcgcctcctggctgagctggagtcc 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1624 GACCAGCAGCTGGATGATCTACAGACCAAGTTTGCTCGCCTCCTGGCTGAGCTGGAGTCC 1683

Qy   1561 agcgcacttaagattgcttaccgcattgaacggctggagtggcagactcgagagtggcca 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1684 AGCGCACTTAAGATTGCTTACCGCATTGAACGGCTGGAGTGGCAGACTCGAGAGTGGCCA 1743

Qy   1621 atgcccgaggacctggctgaggctgatgacgagggtgagcctgaggagggaacttccaaa 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1744 ATGCCCGAGGACCTGGCTGAGGCTGATGACGAGGGTGAGCCTGAGGAGGGAACTTCCAAA 1803

Qy   1681 gatgaagagggcagggccagccaggagggacccccaggtccagagtga 1728
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1804 GATGAAGAGGGCAGGGCCAGCCAGGAGGGACCCCCAGGTCCAGAGTGA 1851
```

RESULT 2
AX350404
LOCUS AX350404 2551 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 29 from Patent WO0181578.
ACCESSION AX350404
VERSION AX350404.1 GI:18616046
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Vernet,C.A., Fernandes,E.R., Gerlach,V., Shimkets,R.A.,
Malyankar,U.M., Boldog,F.L., Zerhusen,B.D., Spytek,K.A.,
Majumder,K., Tchernev,V.T., Padigaru,M., Patturajan,M.,
Burgess,C.E., Gangolli,E.A., Smithson,G., Rastelli,L.,
Macdougall J.R., Taupier,R.J., Grosse,W.M. and Alsobrook,J.P.
TITLE Novel proteins and nucleic acids encoding same
JOURNAL Patent: WO 0181578-A 29 01-NOV-2001;
Curagen Corporation (US)
FEATURES Location/Qualifiers
source 1. .2551
/organism="Homo sapiens"

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 20, 2002, 17:06:57 ; Search time 37.21 Seconds
(without alignments)
598.327 Million cell updates/sec

Title: US-09-927-267-1
Perfect score: 2989
Sequence: 1 MSQDTKVKTTESSPPAPSKA.....EGTSKDEEGRASQEGPPGPE 575

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2811	94.0	575	1	CNGX_RAT	Q64359 rattus norv
2	1579	52.8	706	1	CNG3_BOVIN	Q29441 bos taurus
3	1573.5	52.6	631	1	CNG3_MOUSE	Q9jjz8 mus musculu
4	1568	52.5	663	1	CNG2_BOVIN	Q03041 bos taurus
5	1562.5	52.3	735	1	CNG1_CHICK	Q90805 gallus gall
6	1562	52.3	664	1	CNG2_RABIT	Q28718 oryctolagus
7	1558	52.1	664	1	CNG2_RAT	Q00195 rattus norv
8	1549.5	51.8	694	1	CNG3_HUMAN	Q16281 homo sapien
9	1547	51.8	664	1	CNG2_MOUSE	Q62398 mus musculu
10	1544.5	51.7	690	1	CNG1_BOVIN	Q00194 b cgmp-gate
11	1538.5	51.5	686	1	CNG1_HUMAN	P29973 h cgmp-gate
12	1533.5	51.3	684	1	CNG1_MOUSE	P29974 m cgmp-gate
13	1532.5	51.3	691	1	CNG1_CANFA	Q28279 c cgmp-gate
14	1532	51.3	683	1	CNG1_RAT	Q62927 r cgmp-gate
15	1529	51.2	645	1	CNG3_CHICK	Q90980 gallus gall
16	1436.5	48.1	682	1	CNG_ICTPU	P55934 ictalurus p
17	1235	41.3	665	1	CNG_DROME	Q24278 drosophila
18	1121.5	37.5	741	1	TAX4_CAEEL	Q03611 caenorhabdi
19	847	28.3	261	1	CNG2_HUMAN	Q16280 homo sapien
20	738	24.7	1394	1	CNG4_BOVIN	Q28181 bos taurus
21	719	24.1	909	1	CNG4_HUMAN	Q14028 homo sapien
22	377.5	12.6	1174	1	CIKE_DROME	Q02280 drosophila
23	375	12.5	1159	1	HERG_HUMAN	Q12809 homo sapien
24	159.5	5.3	768	1	KGP1_DROME	Q03042 drosophila
25	133	4.4	377	1	KAPR_APLCA	P31319 aplysia cal
26	129.5	4.3	686	1	KGPB_BOVIN	P21136 bos taurus
27	129.5	4.3	686	1	KGPB_HUMAN	P14619 homo sapien
28	129.5	4.3	686	1	KGPB_MOUSE	Q9z0z0 mus musculu
29	127.5	4.3	379	1	KAP0_BOVIN	P00514 bos taurus
30	126	4.2	327	1	KAPR_DICDI	P05987 dictyosteli
31	125.5	4.2	380	1	KAP0_RAT	P05456 rattus norv
32	125.5	4.2	670	1	KGPA_BOVIN	P00516 bos taurus
33	124.5	4.2	670	1	KGPA_HUMAN	Q13976 homo sapien
34	122.5	4.1	379	1	KAP0_PIG	P07802 sus scrofa
35	121.5	4.1	412	1	KAPR_SCHPO	P36600 schizosacch
36	121.5	4.1	670	1	KGPA_RABIT	O77676 oryctolagus
37	119.5	4.0	486	1	K1C0_XENLA	P02537 xenopus lae
38	119	4.0	390	1	KAPR_MAGGR	O14448 magnaporthe
39	118.5	4.0	762	1	KGP2_HUMAN	Q13237 homo sapien
40	117	3.9	762	1	KGP2_MOUSE	Q61410 mus musculu
41	117	3.9	762	1	KGP2_RAT	Q64595 rattus norv
42	116.5	3.9	381	1	KAP0_HUMAN	P10644 homo sapien
43	116.5	3.9	412	1	KAPR_EMENI	O59922 emericella
44	116.5	3.9	933	1	KGP3_DROME	P32023 drosophila
45	116.5	3.9	1088	1	KGP2_DROME	Q03043 drosophila

ALIGNMENTS

RESULT 1
CNGX_RAT
ID CNGX_RAT STANDARD; PRT; 575 AA.
AC Q64359;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Cyclic-nucleotide-gated olfactory channel OCNC2 subunit.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RX MEDLINE=94377458; PubMed=7522325;
RA Bradley J., Li J., Davidson N., Lester H.A., Zinn K.;
RT "Heteromeric olfactory cyclic nucleotide-gated channels: a subunit
RT that confers increased sensitivity to cAMP.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:8890-8894(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Olfactory neuroepithelium;
RX MEDLINE=95000663; PubMed=7522482;
RA Liman E.R., Buck L.B.;
RT "A second subunit of the olfactory cyclic nucleotide-gated channel
RT confers high sensitivity to cAMP.";
RL Neuron 13:611-621(1994).
RN [3]
RP SEQUENCE OF 7-35 FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RA Bradley J., Zhang Y., Bakin R., Lester H.A., Ronnett G., Zinn K.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: ODORANT SIGNAL TRANSDUCTION IS PROBABLY MEDIATED
CC BY A G-PROTEIN COUPLED CASCADE USING CAMP AS SECOND MESSENGER.
CC THE OLFACTORY CHANNEL CAN BE SHOWN TO BE ACTIVATED BY CYCLIC
CC NUCLEOTIDES WHICH LEADS TO A DEPOLARIZATION OF OLFACTORY
CC SENSORY NEURONS.
CC -!- SUBUNIT: HETEROOLIGOMER OF OCNC1 AND OCNC2 SUBUNITS.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: OLFACTORY NEURONS.
CC ----------
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ----------
DR EMBL; U12623; AAA21464.1; -.
DR EMBL; U12425; AAA64748.1; -.
DR EMBL; U76219; AAC17596.1; -.
DR InterPro; IPR000636; Cation_chan_non_lig.
DR InterPro; IPR000595; cNMP_binding.
DR Pfam; PF00027; cNMP_binding; 1.

THIS PAGE BLANK (USPTO)

```
DR   Pfam; PF00520; ion_trans; 1.
DR   SMART; SM00100; cNMP; 1.
DR   PROSITE; PS00888; CNMP_BINDING_1; 1.
DR   PROSITE; PS00889; CNMP_BINDING_2; 1.
DR   PROSITE; PS50042; CNMP_BINDING_3; 1.
KW   Ionic channel; Ion transport; cAMP-binding; Transmembrane;
KW   Multigene family; Olfaction.
FT   DOMAIN        1     33       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     34     54       H1 (POTENTIAL).
FT   DOMAIN       55     65       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     66     86       H2 (POTENTIAL).
FT   DOMAIN       87    112       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    113    133       H3 (POTENTIAL).
FT   DOMAIN      134    168       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    169    189       H4 (POTENTIAL).
FT   DOMAIN      190    244       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    245    265       H5 (POTENTIAL).
FT   DOMAIN      266    400       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    401    421       H6 (POTENTIAL).
FT   DOMAIN      422    575       CYTOPLASMIC (POTENTIAL).
FT   NP_BIND     356    500       CAMP (BY SIMILARITY).
FT   BINDING     415    415       CAMP (POTENTIAL).
FT   BINDING     430    430       CAMP (POTENTIAL).
SQ   SEQUENCE   575 AA;  65674 MW;  FCF01C563214F575 CRC64;


  Query Match           94.0%;  Score 2811;  DB 1;  Length 575;
  Best Local Similarity   93.4%;  Pred. No. 4e-186;
  Matches  537;  Conservative   16;  Mismatches   22;  Indels    0;  Gaps    0;

Qy       1 MSQDTKVKTTESSPPAPSKARKLLPVLDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPD 60
           |||| |||||||:||||:|||| |||||||||||||||||||||:||||||:||||||||
Db       1 MSQDGKVKTTESTPPAPTKARKWLPVLDPSGDYYYWWLNTMVFPIMYNLIIVVCRACFPD 60

Qy      61 LQHGYLVAWLVLDYTSDLLYLLDMVVRFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLA 120
           ||| ||||| |||||||||||||: |||||||||||||||||| |:|||||||| ||||
Db      61 LQHSYLVAWFVLDYTSDLLYLLDIGVRFHTGFLEQGILVVDKGMIASRYVRTWSFLLDLA 120

Qy     121 SLMPTDVVYVRLGPHTPTLRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVI 180
           ||:|||  ||:|||| |||||||||| ||||||||||||||||||||||||||||||||
Db     121 SLVPTDAAYVQLGPHIPTLRLNRFLRVPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVI 180

Qy     181 HWNSCLYFALSRYLGFGRDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAR 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||| | |
Db     181 HWNSCLYFALSRYLGFGRDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPLPDR 240

Qy     241 EEEYLFMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLE 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||::||
Db     241 EEEYLFMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNKRLE 300

Qy     301 RRVIDWYQHLQINKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEEL 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 RRVIDWYQHLQINKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEEL 360

Qy     361 VLKLQPQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIIN 420
           ||||||||||||||||||:||||||||||||||:||||||||||||||||||||||||||
Db     361 VLKLQPQTYSPGEYVCRKGDIGREMYIIREGQLAVVADDGVTQYAVLGAGLYFGEISIIN 420

Qy     421 IKGNMSGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLDV 480
           |||||||||||||||||||||||||||||||||||||||| :|||||||||||||||||
Db     421 IKGNMSGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQAQAVMEEKGREILLKMNKLDV 480

Qy     481 NAEAAEIALQEATESRLRGLDQQLDDLQTKFARLLAELESSALKIAYRIERLEWQTREWP 540
           |||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||
Db     481 NAEAAEIALQEATESRLKGLDQQLDDLQTKFARLLAELESSALKIAYRIERLEWQTREWP 540

Qy     541 MPEDLAEADDEGEPEEGTSKDEEGRASQEGPPGPE 575
           ||||: ||||| || |||||| ||:| | || | |
Db     541 MPEDMGEADDEAEPGEGTSKDGEGKAGQAGPSGIE 575


RESULT   2
CNG3_BOVIN
ID   CNG3_BOVIN     STANDARD;      PRT;   706 AA.
AC   Q29441;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Cyclic-nucleotide-gated cation channel alpha 3 (CNG channel alpha 3)
DE   (CNG-3) (CNG3) (Cyclic nucleotide gated channel alpha 3) (Cone
DE   photoreceptor cGMP-gated channel alpha subunit).
GN   CNGA3 OR CNCG3.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RX   MEDLINE=94211295; PubMed=7512693;
RA   Weyand I., Godde M., Frings S., Weiner J., Mueller F., Altenhofen W.,
RA   Hatt H., Kaupp U.B.;
RT   "Cloning and functional expression of a cyclic-nucleotide-gated
RT   channel from mammalian sperm.";
RL   Nature 368:859-863(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RX   MEDLINE=94224768; PubMed=8170936;
RA   Biel M., Zong X., Distler M., Bosse E., Klugbauer N., Murakami M.,
RA   Flockerzi V., Hofmann F.;
RT   "Another member of the cyclic nucleotide-gated channel family,
RT   expressed in testis, kidney, and heart.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:3505-3509(1994).
CC   -!- FUNCTION: COULD BE RESPONSIBLE FOR CGMP-INDUCED CALCIUM ENTRY IN
CC       CELLS OTHER THAN SENSORY CELLS. MIGHT BE INVOLVED IN CHEMOTAXIS OF
CC       SPERM.
CC   -!- SUBUNIT: FORMS FUNCTIONAL HETEROOLIGOMERIC CHANNELS WITH CNG4 IN
CC       VITRO.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: TESTIS, KIDNEY, RETINAL CONE, AND HEART.
CC   -!- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE-GATED CATION CHANNEL
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X89600; CAA61759.1; -.
DR   EMBL; X76485; CAA54023.1; -.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR000595; cNMP_binding.
DR   Pfam; PF00027; cNMP_binding; 1.
DR   Pfam; PF00520; ion_trans; 1.
DR   SMART; SM00100; cNMP; 1.
DR   PROSITE; PS00888; CNMP_BINDING_1; 1.
DR   PROSITE; PS00889; CNMP_BINDING_2; 1.
DR   PROSITE; PS50042; CNMP_BINDING_3; 1.
KW   Ionic channel; Ion transport; cAMP-binding; Transmembrane;
KW   Multigene family; Vision.
FT   TRANSMEM    186    206       POTENTIAL.
FT   TRANSMEM    322    342       POTENTIAL.
FT   TRANSMEM    398    418       POTENTIAL.
FT   NP_BIND     501    624       CAMP.
FT   BINDING     568    568       CAMP (POTENTIAL).
FT   BINDING     583    583       CAMP (POTENTIAL).
SQ   SEQUENCE   706 AA;  81132 MW;  F4990DCD29B56239 CRC64;


  Query Match           52.8%;  Score 1579;  DB 1;  Length 706;
  Best Local Similarity   52.7%;  Pred. No. 3.4e-101;
```

THIS PAGE BLANK (USPTO)

```
    1464 tGluGluLysGlyArgGluIleLeuLeuLysMetAsnLysLeuAspValA 481
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1504 GGAGGAGAAAGGACGTGAGATCCTGCTGAAAATGAACAAGTTGGACGTGA 1503

     481 sqAlaGluAlaAlaGluIleAlaLeuGlnGluAlaThrGluSerArgLeu 497
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1504 ATGCTGAGGCAGCTGAGATCGCCCTGCAGGAGGCCACAGAGTCCCGGCTA 1553

     498 ArgGlyLeuAspGlnGlnLeuAspAspLeuGlnThrLysPheAlaArgLe 514
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1554 CGAGGCCTAGACCAGCAGCTGGATGATCTACAGACCAAGTTTGCTCGCCT 1603

     514 uLeuAlaGluLeuGluSerSerAlaLeuLysIleAlaTyrArgIleGluA 531
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1604 CCTGGCTGAGCTGGAGTCCAGCGCACTTAAGATTGCTTACCGCATTGAAC 1653

     531 rgLeuGluTrpGlnThrArgGluTrpProMetProGluAspLeuAlaGlu 547
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1654 GGCTGGAGTGGCAGACTCGAGACTGGCCAATGCCCGAGGACCTGGCTGAG 1703

     548 AlaAspAspGluGlyGluProGluGluGlyThrSerLysAspGluGluGl 564
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1704 GCTGATGACGAGGGTGAGCCTGAGGAGGGAACTTCCAAAGATGAAGAGGG 1753

      564 yArgAlaSerGlnGluGlyProProGlyProGlu 575
          ||||||||||||||||||||||||||||||||||
     1754 CAGGGCCAGCCAGGAGGGACCCCCAGGTCCAGAG 1787

seq_name: gb_ro:RNU12425

seq_documentation_block:
LOCUS       RNU12425                1938 bp    mRNA    linear   ROD 05-APR-1995
DEFINITION  Rattus norvegicus olfactory cyclic nucleotide-gated channel mRNA,
            complete cds.
ACCESSION   U12425
VERSION     U12425.1  GI:548083
KEYWORDS    .
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1938)
  AUTHORS   Liman,E.R. and Buck,L.B.
  TITLE     A second subunit of the olfactory cyclic nucleotide-gated channel
            confers high sensitivity to cAMP
  JOURNAL   Neuron 13 (3), 611-621 (1994)
  MEDLINE   95000663
REFERENCE   2  (bases 1 to 1938)
  AUTHORS   Liman,E.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-1994) Emily R. Liman, Neurobiology, Harvard
            Medical School, 220 Longwood Avenue, Boston, MA 02115, USA
FEATURES             Location/Qualifiers
     source          1. .1938
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /tissue_type="olfactory neuroepithelium"
     mRNA            1. .1937
                     /evidence=experimental
     CDS             107. .1834
                     /codon_start=1
                     /product="olfactory cyclic nucleotide-gated channel"
                     /protein_id="AAA64748.1"
                     /db_xref="GI:548084"
                     /translation="MSQDGKVKTTESTPPAPTKARKWLPVLDPSGDYYYWWLNTMVFP
                     IMYNLIIVVCRACFPDLQHSYLVAWFVLDYTSDLLYLLDIGVRFHTGFLEQGILVVDK
                     GMIASRYVRTWSFLLDLASLVPTDAAYVQLGPHIPTLRLNRFLRVPRLFEAFDRTETR
                     TAYPNAFRIAKLMLYIFVVIHWNSCLYFALSRYLGFGRDAWVYPDPAQPGFERLRRQY
                     LYSFYFSTLILTTVGDTPLPDREEEYLFMVGDFLLAVMGFATIMGSMSSVIYNMNTAD
                     AAFYPDHALVKKYMKLQHVNKRLERRVIDWYQHLQINKKMTNEVAILQHLPERLRAEV
                     AVSVHLSTLSRVQIFQNCEASLLEELVLKLQPQTYSPGEYVCRKGDIGREMYIIREGQ
                     LAVVADDGVTQYAVLGAGLYFGEISIINIKGNMSGNRRTANIKSLGYSDLFCLSKEDL
                     REVLSEYPQAQAVMEEKGREILLKMNKLDVNAEAAEIALQEATESRLKGLDQQLDDLQ
                     TKFARLLAELESSALKIAYRIERLEWQTREWPMPEDMGEADDEAEPGEGTSKDGEGKA
                     GQAGPSGIE"
     polyA_site      1938
                     /note="3 A residues"
BASE COUNT      458 a    531 c    543 g    406 t
ORIGIN


alignment_scores:
           Quality: 2811.00              Length:    575
             Ratio:  4.993                Gaps:      0
 Percent Similarity:  97.913   Percent Identity: 93.391

alignment_block:
 US-09-927-267-1 x RNU12425   ..

 Align seg 1/1  to: RNU12425  from: 1  to: 1938

       1 MetSerGlnAspThrLysValLysThrThrGluSerSerProProAlaPr 17
         ||||||||||||   ||||||||||||||||||||||:::||||||||||||
     107 ATGAGCCAGGACGGCAAAGTGAAGACCACAGAGTCCACGCCCCCAGCCCC 156

      17 oSerLysAlaArgLysLeuLeuProValLeuAspProSerGlyAspTyrT 34
         |:::||||||||||||   ||||||||||||||||||||||||||||||.
     157 AACCAAAGCCAGGAAGTGGCTGCCAGTCCTAGACCCATCTGGGGATTACT 206

      34 yrTyrTrpTrpLeuAsnThrMetValPheProValMetTyrAsnLeuIle 50
         ||||||||||||||||||||||||||||||:::||||||||||||||||
     207 ACTACTGGTGGCTGAACACAATGGTCTTCCCAATCATGTACAACCTCATC 256

      51 IleLeuValCysArgAlaCysPheProAspLeuGlnHisGlyTyrLeuVa 67
         |||:::|||||||||||||||||||||||||||||||||:::||||||||
     257 ATCGTTGTATGCAGGGCCTGCTTTCCTGACTTGCAGCACAGTTACCTGGT 306

      67 lAlaTrpLeuValLeuAspTyrThrSerAspLeuLeuTyrLeuLeuAspM 84
         |||||||:::|||||||||||||||||||||||||||||||||||||||:
     307 GGCCTGGTTCGTGCTGGACTACACGAGTGACCTGCTGTACCTACTAGATA 356

      84 etValValArgPheHisThrGlyPheLeuGluGlnGlyIleLeuValVal 100
         ::   ||||||||||||||||||||||||||||||||||||||||||||||
     357 TTGGGGTACGCTTCCACACAGGATTCCTAGAGCAGGGCATCCTGGTGGTA 406

     101 AspLysGlyArgIleSerSerArgTyrValArgThrTrpSerPhePheLe 117
         |||||||||   |||:::|||||||||||||||||||||||||||:::||
     407 GACAAAGGCATGATCGCCAGTCGCTACGTCCGCACCTGGAGCTTCCTGTT 456

     117 uAspLeuAlaSerLeuMetProThrAspValValTyrValArgLeuGlyP 134
         |||||||||||||||:::|||||||||:::::|||||||:::|||||||
     457 GGACCTGGCGTCCCTGGTCCCCACAGATGCGGCCTATGTGCAGCTGGGCC 506

     134 roHisThrProThrLeuArgLeuAsnArgPheLeuArgAlaProArgLeu 150
         |||||   |||||||||||||||||||||||||||:::|||||||||||
     507 CCCACATCCCTACACTCCGGCTAAACCGCTTTCTCCGAGTGCCCCGCCTC 556

     151 PheGluAlaPheAspArgThrGluThrArgThrAlaTyrProAsnAlaPh 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     557 TTCGAGGCTTTTGATCGTACAGAGACCCGCACGGCTTACCCAAATGCCTT 606

     167 eArgIleAlaLysLeuMetLeuTyrIlePheValValIleHisTrpAsnS 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     607 CCGCATAGCCAAGCTGATGCTTTACATTTTTGTTGTCATCCATTGGAACA 656

     184 erCysLeuTyrPheAlaLeuSerArgTyrLeuGlyPheGlyArgAspAla 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     657 GTTGCTTATACTTCGCCCTGTCCAGGTACCTGGGCTTTGGACGGGATGCG 706

     201 TrpValTyrProAspProAlaGlnProGlyPheGluArgLeuArgArgGl 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     707 TGGGTATACCCAGACCCTGCGCAACCTGGCTTTGAGCGCTTGCGGCGCCA 756

     217 nTyrLeuTyrSerPheTyrPheSerThrLeuIleLeuThrThrValGlyA 234
```

THIS PAGE BLANK (USPTO)

```
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 757 GTATCTCTACAGCTTCTACTTCTCCACTCTGATCCTGACCACAGTGGGTG 806

 234 spThrProProProAlaArgGluGluGluTyrLeuPheMetValGlyAsp 250
     ||||||||   |||  ||||||||||||||||||||||||||||||||||
 807 ACACGCCGCTGCCAGACCGAGAGGAAGAGTACCTCTTCATGGTGGGTGAC 856

 251 PheLeuLeuAlaValMetGlyPheAlaThrIleMetGlySerMetSerSe 267
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 857 TTCCTGCTGGCCGTCATGGGTTTCGCCACCATCATGGGTAGCATGAGCTC 906

 267 rValIleTyrAsnMetAsnThrAlaAspAlaAlaPheTyrProAspHisA 284
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 907 TGTCATTTACAACATGAACACTGCAGATGCGGCCTTCTACCCAGACCATG 956

 284 laLeuValLysLysTyrMetLysLeuGlnHisValAsnArgLysLeuGlu 300
     ||||||||||||||||||||||||||||||||||||||::::::||||||
 957 CGCTGGTAAAGAAGTACATGAAGCTGCAGCATGTCAACAAGAGGCTGGAG 1006

 301 ArgArgValIleAspTrpTyrGlnHisLeuGlnIleAsnLysLysMetTh 317
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1007 CGGCGAGTTATTGACTGGTACCAGCATCTTCAGATCAACAAGAAGATGAC 1056

 317 rAsnGluValAlaIleLeuGlnHisLeuProGluArgLeuArgAlaGluV 334
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1057 CAACGAGGTAGCCATCTTGCAGCACCTGCCTGAGCGGCTGCGGGCGGAGG 1106

 334 alAlaValSerValHisLeuSerThrLeuSerArgValGlnIlePheGln 350
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1107 TTGCTGTGTCCGTGCACCTGTCTACCCTGAGCCGAGTACAGATCTTCCAG 1156

 351 AsnCysGluAlaSerLeuLeuGluGluLeuValLeuLysLeuGlnProGl 367
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1157 AACTGTGAAGCCAGCCTGCTGGAAGAGCTGGTGCTGAAGCTACAGCCCCA 1206

 367 nThrTyrSerProGlyGluTyrValCysArgLysGlyAspIleGlyGlnG 384
     ||||||||||||||||||||||||||||||||||||||||||||||:::|
1207 GACCTACTCGCCAGGCGAATATGTGTGCCGCAAAGGGGACATTGGCCGAG 1256

 384 luMetTyrIleIleArgGluGlyGlnLeuAlaValValAlaAspAspGly 400
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1257 AGATGTACATCATCCGTGAGGGCCAGCTGGCTGTGGTGGCCGATGATGGT 1306

 401 IleThrGlnTyrAlaValLeuGlyAlaGlyLeuTyrPheGlyGluIleSe 417
     :::||||||||||||||||||||||||||||||||||||||||||||||||
1307 GTCACACAGTATGCTGTGCTTGGTGCAGGGCTCTACTTTGGGGAGATCAG 1356

 417 rIleIleAsnIleLysGlyAsnMetSerGlyAsnArgArgThrAlaAsnI 434
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1357 TATCATCAACATCAAAGGGAACATGTCTGGAAACCGACGAACAGCCAACA 1406

 434 leLysSerLeuGlyTyrSerAspLeuPheCysLeuSerLysGluAspLeu 450
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1407 TCAAGAGCCTAGGTTATTCAGACCTGTTCTGCCTCAGCAAGGAGGATCTG 1456

 451 ArgGluValLeuSerGluTyrProGlnAlaGlnThrIleMetGluGluLy 467
     ||||||||||||||||||||||||||||||||||:::::||||||||||||
1457 CGGGAGGTACTGAGTGAGTACCCACAGGCCCAGGCGGTCATGGAGGAGAA 1506

 467 sGlyArgGluIleLeuLeuLysMetAsnLysLeuAspValAsnAlaGluA 484
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1507 GGGCCGAGAAATCTTGCTCAAAATGAATAAGTTGGATGTGAATGCTGAGG 1556

 484 laAlaGluIleAlaLeuGlnGluAlaThrGluSerArgLeuArgGlyLeu 500
     ||||||||||||||||||||||||||||||||||||||:::|||||||||
1557 CAGCTGAGATCGCCCTCCAGGAGGCCACAGAGTCTCGGCTCAAAGGCCTC 1606

 501 AspGlnGlnLeuAspAspLeuGlnThrLysPheAlaArgLeuLeuAlaGl 517
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1607 GACCAGCAGCTTGATGATCTGCAGACCAAGTTTGCTCGCCTACTGGCTGA 1656

 517 uLeuGluSerSerAlaLeuLysIleAlaTyrArgIleGluArgLeuGluT 534
     ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
1657 GCTGGAGTCCAGTGCACTGAAGATAGCTTACCGCATCGAAAGGCTGGAGT 1706

 534 rpGlnThrArgGluTrpProMetProGluAspLeuAlaGluAlaAspAsp 550
     ||||||||||||||||||||||||||||||||::::::|||||||||||||
1707 GGCAGACTCGAGAGTGGCCAATGCCAGAGGACATGGGTGAGGCTGATGAT 1756

 551 GluGlyGluProGluGluGlyThrSerLysAspGluGluGlyArgAlaSe 567
     |||:::||||||   ||||||||||||||||||   |||||||:::|||::
1757 GAGGCTGAGCCTGGAGAAGGGACGTCCAAGGATGGAGAGGGAAAGGCTGG 1806

 567 rGlnGluGlyProProGlyProGlu 575
     :|||   ||||||   |||   |||
1807 CCAGGCGGGACCCTCAGGCATAGAA 1831
```

```
seq_name: gb_ro:RNU12623

seq_documentation_block:
LOCUS       RNU12623                2351 bp    mRNA    linear   ROD 15-SEP-1994
DEFINITION  Rattus norvegicus Sprague-Dawley cyclic nucleotide gated cation
            channel mRNA, complete cds.
ACCESSION   U12623
VERSION     U12623.1  GI:538128
KEYWORDS    .
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 2351)
  AUTHORS   Bradley,J., Li,J., Davidson,N., Lester,H.A. and Zinn,K.
  TITLE     Heteromeric olfactory cyclic nucleotide-gated channels: A new
            subunit that confers increased sensitivity to cAMP
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 91, 8890-8894 (1994)
  MEDLINE   94377458
REFERENCE   2  (bases 1 to 2351)
  AUTHORS   Bradley,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-1994) Jonathan Bradley, Biology, California
            Institute of Technology, Pasadena, CA 91125, USA
COMMENT     On Sep 15, 1994 this sequence version replaced gi:521147.
FEATURES             Location/Qualifiers
     source          1. .2351
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
     CDS             109. .1836
                     /codon_start=1
                     /product="cyclic nucleotide gated cation channel"
                     /protein_id="AAA21464.1"
                     /db_xref="GI:538129"
                     /translation="MSQDGKVKTTESTPPAPTKARKWLPVLDPSGDYYYWWLNTMVFP
                     IMYNLIIVVCRACFPDLQHSYLVAWFVLDYTSDLLYLLDIGVRFHTGFLEQGILVVDK
                     GMIASRYVRTWSFLLDLASLVPTDAAYVQLGPHIPTLRLNRFLRVPRLFEAFDRTETR
                     TAYPNAFRIAKLMLYIFVVIHWNSCLYFALSRYLGFGRDAWVYPDPAQPGFERLRRQY
                     LYSFYFSTLILTTVGDTPLPDREEEYLFMVGDFLLAVMGFATIMGSMSSVIYNMNTAD
                     AAFYPDHALVKKYMKLQHVNKRLERRVIDWYQHLQINKKMTNEVAILQHLPERLRAEV
                     AVSVHLSTLSRVQIFQNCEASLLEELVLKLQPQTYSPGEYVCRKGDIGREMYIIREGQ
                     LAVVADDGVTQYAVLGAGLYFGEISIINIKGNMSGNRRTANIKSLGYSDLFCLSKEDL
                     REVLSEYPQAQAVMEEKGREILLKMNKLDVNAEAAEIALQEATESRLKGLDQQLDDLQ
                     TKFARLLAELESSALKIAYRIERLEWQTREWPMPEDMGEADDEAEPGEGTSKDGEGKA
                     GQAGPSGIE"
BASE COUNT      580 a    624 c    637 g    510 t
ORIGIN


alignment_scores:
            Quality: 2811.00              Length:    575
              Ratio:  4.993                Gaps:      0
 Percent Similarity:  97.913   Percent Identity: 93.391

alignment_block:
 US-09-927-267-1 x RNU12623    ..
```

THIS PAGE BLANK (USPTO)

```
Align seg 1/1  to: RNU12623  from: 1  to: 2351

      1 MetSerGlnAspThrLysValLysThrThrGluSerSerProProAlaPr 17
    109 ATGAGCCAGGACGGCAAAGTGAAGACCACAGAGTCCACGCCCCCAGCCCC 158

     17 oSerLysAlaArgLysLeuLeuProValLeuAspProSerGlyAspTyrT 34
    159 AACCAAAGCCAGGAAGTGGCTGCCAGTCCTAGACCCATCTGGGGATTACT 208

     34 yrTyrTrpTrpLeuAsnThrMetValPheProValMetTyrAsnLeuIle 50
    209 ACTACTGGTGGCTGAACACAATGGTCTTCCCAATCATGTACAACCTCATC 258

     51 IleLeuValCysArgAlaCysPheProAspLeuGlnHisGlyTyrLeuVa 67
    259 ATCGTTGTATGCAGGGCCTGCTTTCCTGACTTGCAGCACAGTTACCTGGT 308

     67 lAlaTrpLeuValLeuAspTyrThrSerAspLeuLeuTyrLeuLeuAspM 84
    309 GGCCTGGTTCGTGCTGGACTACACGAGTGACCTGCTGTACCTACTAGATA 358

     84 etValValArgPheHisThrGlyPheLeuGluGlnGlyIleLeuValVal 100
    359 TTGGGGTACGCTTCCACACAGGATTCCTAGAGCAGGGCATCCTGGTGGTA 408

    101 AspLysGlyArgIleSerSerArgTyrValArgThrTrpSerPhePheLe 117
    409 GACAAAGGCATGATCGCCAGTCGCTACGTCCGCACCTGGAGCTTCCTGTT 458

    117 uAspLeuAlaSerLeuMetProThrAspValValTyrValArgLeuGlyP 134
    459 GGACCTGGCGTCCCTGGTCCCCACAGATGCGGCCTATGTGCAGCTGGGCC 508

    134 roHisThrProThrLeuArgLeuAsnArgPheLeuArgAlaProArgLeu 150
    509 CCCACATCCCTACACTCCGGCTAAACCGCTTTCTCCGAGTGCCCCGCCTC 558

    151 PheGluAlaPheAspArgThrGluThrArgThrAlaTyrProAsnAlaPh 167
    559 TTCGAGGCTTTTGATCGTACAGAGACCCGCACGGCTTACCCAAATGCCTT 608

    167 eArgIleAlaLysLeuMetLeuTyrIlePheValValIleHisTrpAsnS 184
    609 CCGCATAGCCAAGCTGATGCTTTACATTTTTGTTGTCATCCATTGGAACA 658

    184 erCysLeuTyrPheAlaLeuSerArgTyrLeuGlyPheGlyArgAspAla 200
    659 GTTGCTTATACTTCGCCCTGTCCAGGTACCTGGGCTTTGGACGGGATGCG 708

    201 TrpValTyrProAspProAlaGlnProGlyPheGluArgLeuArgArgGl 217
    709 TGGGTATACCCAGACCCTGCGCAACCTGGCTTTGAGCGCTTGCGGCGCCA 758

    217 nTyrLeuTyrSerPheTyrPheSerThrLeuIleLeuThrThrValGlyA 234
    759 GTATCTCTACAGCTTCTACTTCTCCACTCTGATCCTGACCACAGTGGGTG 808

    234 spThrProProProAlaArgGluGluGluTyrLeuPheMetValGlyAsp 250
    809 ACACGCCGCTGCCAGACCGAGAGGAAGAGTACCTCTTCATGGTGGGTGAC 858

    251 PheLeuLeuAlaValMetGlyPheAlaThrIleMetGlySerMetSerSe 267
    859 TTCCTGCTGGCCGTCATGGGTTTCGCCACCATCATGGGTAGCATGAGCTC 908

    267 rValIleTyrAsnMetAsnThrAlaAspAlaAlaPheTyrProAspHisA 284
    909 TGTCATTTACAACATGAACACTGCAGATGCGGCCTTCTACCCAGACCATG 958

    284 laLeuValLysLysTyrMetLysLeuGlnHisValAsnArgLysLeuGlu 300
    959 CGCTGGTAAAGAAGTACATGAAGCTGCAGCATGTCAACAAGAGGCTGGAG 1008

    301 ArgArgValIleAspTrpTyrGlnHisLeuGlnIleAsnLysLysMetTh 317
   1009 CGGCGAGTTATTGACTGGTACCAGCATCTTCAGATCAACAAGAAGATGAC 1058

    317 rAsnGluValAlaIleLeuGlnHisLeuProGluArgLeuArgAlaGluV 334
   1059 CAACGAGGTAGCCATCTTGCAGCACCTGCCTGAGCGGCTGCGGGCGGAGG 1108

    334 alAlaValSerValHisLeuSerThrLeuSerArgValGlnIlePheGln 350
   1109 TTGCTGTGTCCGTGCACCTGTCTACCCTGAGCCGAGTACAGATCTTCCAG 1158

    351 AsnCysGluAlaSerLeuLeuGluGluLeuValLeuLysLeuGlnProGl 367
   1159 AACTGTGAAGCCAGCCTGCTGGAAGAGCTGGTGCTGAAGCTACAGCCCCA 1208

    367 nThrTyrSerProGlyGluTyrValCysArgLysGlyAspIleGlyGlnG 384
   1209 GACCTACTCGCCAGGCGAATATGTGTGCCGCAAAGGGGACATTGGCCGAG 1258

    384 luMetTyrIleIleArgGluGlyGlnLeuAlaValValAlaAspAspGly 400
   1259 AGATGTACATCATCCGTGAGGGCCAGCTGGCTGTGGTGGCCGATGATGGT 1308

    401 IleThrGlnTyrAlaValLeuGlyAlaGlyLeuTyrPheGlyGluIleSe 417
   1309 GTCACACAGTATGCTGTGCTTGGTGCAGGGCTCTACTTTGGGGAGATCAG 1358

    417 rIleIleAsnIleLysGlyAsnMetSerGlyAsnArgArgThrAlaAsnI 434
   1359 TATCATCAACATCAAAGGGAACATGTCTGGAAACCGACGAACAGCCAACA 1408

    434 leLysSerLeuGlyTyrSerAspLeuPheCysLeuSerLysGluAspLeu 450
   1409 TCAAGAGCCTAGGTTATTCAGACCTGTTCTGCCTCAGCAAGGAGGATCTG 1458

    451 ArgGluValLeuSerGluTyrProGlnAlaGlnThrIleMetGluGluLy 467
   1459 CGGGAGGTACTGAGTGAGTACCCACAGGCCCAGGCGGTCATGGAGGAGAA 1508

    467 sGlyArgGluIleLeuLeuLysMetAsnLysLeuAspValAsnAlaGluA 484
   1509 GGGCCGAGAAATCTTGCTCAAAATGAATAAGTTGGATGTGAATGCTGAGG 1558

    484 laAlaGluIleAlaLeuGlnGluAlaThrGluSerArgLeuArgGlyLeu 500
   1559 CAGCTGAGATCGCCCTCCAGGAGGCCACAGAGTCTCGGCTCAAAGGCCTC 1608

    501 AspGlnGlnLeuAspAspLeuGlnThrLysPheAlaArgLeuLeuAlaGl 517
   1609 GACCAGCAGCTTGATGATCTGCAGACCAAGTTTGCTCGCCTACTGGCTGA 1658

    517 uLeuGluSerSerAlaLeuLysIleAlaTyrArgIleGluArgLeuGluT 534
   1659 GCTGGAGTCCAGTGCACTGAAGATAGCTTACCGCATCGAAAGGCTGGAGT 1708

    534 rpGlnThrArgGluTrpProMetProGluAspLeuAlaGluAlaAspAsp 550
   1709 GGCAGACTCGAGAGTGGCCAATGCCAGAGGACATGGGTGAGGCTGATGAT 1758

    551 GluGlyGluProGluGluGlyThrSerLysAspGluGluGlyArgAlaSe 567
   1759 GAGGCTGAGCCTGGAGAAGGGACGTCCAAGGATGGAGAGGGAAAGGCTGG 1808

    567 rGlnGluGlyProProGlyProGlu 575
   1809 CCAGGCGGGACCCTCAGGCATAGAA 1833

seq_name: gb_htg:AC021935

seq_documentation_block:
LOCUS       AC021935              162965 bp    DNA     linear   HTG 03-MAR-2001
```

THIS PAGE BLANK (USPTO)

OM of: US-09-927-267-1 to: N_Geneseq_032802:* out_format : pfs

Date: Jun 20, 2002 6:35 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1p
-Q=/cgn2_1/USPTO_spool/US09927267/runat_19062002_090143_25524/app_query.fasta_1.640
-DB=N_Geneseq_032802 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09927267_@CGN1_1_663 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-927-267-1
Query length: 575
Database: N_Geneseq_032802:*
Database sequences: 1736436
Database length: 858457221
Search time (sec): 267.690000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	!	Documentation	..	
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:ABA04603	+	2984.00	5715.50	0.0					25
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:ABA04602	+	2957.50	5668.28	0.0					18
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAD09558	+	1535.00	2923.86	1.6e-154					1
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL13353	+	1235.00	2340.84	4.9e-122					2
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH98838	+	959.50	1822.14	3.8e-93					9
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL18085	+	947.00	1789.19	2.6e-91					2
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL14931	+	867.50	1625.39	3.5e-82					5
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL13352	+	863.50	1611.61	2.0e-81					8
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL19457	+	781.50	1464.82	3.0e-73					3
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS91658	+	714.00	1332.86	6.8e-66					3
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL19456	-	695.00	1291.49	1.4e-63					5
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:AAD25729	+	655.00	1222.88	9.1e-60					27
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS91258	+	571.00	1071.44	2.5e-51					1
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS79945	+	540.00	1011.43	5.5e-48					1
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS77668	+	540.00	1002.62	1.7e-47					2
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL06311	+	503.00	938.76	6.1e-44					1
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL18084	-	502.00	906.28	3.9e-42					1
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAZ09486	+	480.50	884.64	6.3e-41					31
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL14930	-	478.00	863.46	9.5e-40					1
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAZ09496	+	467.50	854.86	2.9e-39					47
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH48730	+	467.50	854.10	3.2e-39					5
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:ABA09197	+	467.50	853.14	3.6e-39					5
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAX84444	+	466.50	865.87	7.0e-40					15
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH48729	+	465.50	855.04	2.8e-39					3
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH48731	+	464.50	854.09	3.2e-39					3
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC66779	+	456.50	842.06	1.5e-38					23
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAZ09488	+	451.00	834.34	4.0e-38					18
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAZ09485	+	445.00	816.94	3.7e-37					29
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:AAS18223	+	441.50	815.68	4.4e-37					18
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:AAS18224	+	441.50	815.68	4.4e-37					18
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAX84443	+	441.50	813.45	5.8e-37					22
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAX84442	+	441.50	811.23	7.7e-37					27
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:AAS18213	+	441.50	810.98	8.0e-37					27
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:AAS18215	+	441.50	810.98	8.0e-37					27
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:AAS18214	+	441.50	810.21	8.8e-37					29
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:AAS18216	+	441.50	810.21	8.8e-37					29
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAZ09493	+	441.50	810.19	8.8e-37					29
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAZ09484	+	435.50	800.22	3.2e-36					26
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAZ09491	+	432.00	792.26	8.8e-36					28
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAZ09487	+	430.50	789.23	1.3e-35					29
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL14239	+	422.00	769.36	1.7e-34					
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAX84445	+	403.50	742.93	4.9e-33					
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH98302	+	400.50	729.48	2.8e-32					
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAA62050	+	400.50	729.45	2.8e-32					
/SIDS5/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAZ35716	+	395.50	721.87	7.3e-32					

seq_name: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:ABA04603

seq_documentation_block:
ID ABA04603 standard; cDNA; 2551 BP.
XX
AC ABA04603;
XX
DT 21-FEB-2002 (first entry)
XX
DE MOL10b coding sequence.
XX
KW MOL; G-coupled protein-receptor; cardiomyopathy; atherosclerosis;
KW cell signal processing; metabolic disorder; diabetes; cancer; human;
KW neurodegenerative disorder; immune disorder; cardiac disorder;
KW lung disease; autoimmune disease; developmental disorder; antidiabetic;
KW Cytostatic; Neuroprotective; Antiatherosclerotic; Immunosuppressive;
KW Gene therapy; Vaccine; antiinflammatory; MOL10b; chromosome 11; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 779..2506
FT /*tag= a
FT /product= "MOL 10b"
XX
PN WO200181578-A2.
XX
PD 01-NOV-2001.
XX
PF 26-APR-2001; 2001WO-US13578.
XX
PR 26-APR-2000; 2000US-200158P.
PR 28-APR-2000; 2000US-200613P.
PR 28-APR-2000; 2000US-200780P.
PR 01-MAY-2000; 2000US-201006P.
PR 01-MAY-2000; 2000US-201007P.
PR 01-MAY-2000; 2000US-201236P.
PR 01-MAY-2000; 2000US-201238P.
PR 02-MAY-2000; 2000US-201186P.
PR 03-MAY-2000; 2000US-201474P.
PR 03-MAY-2000; 2000US-201508P.
PR 25-JUL-2000; 2000US-220591P.
PR 15-SEP-2000; 2000US-232678P.
PR 22-JAN-2001; 2001US-263217P.
PR 30-JAN-2001; 2001US-265160P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Vernet CAM, Fernandes ER, Gerlach V, Shimkets RA, Malyankar UM;
PI Boldog FL, Zerhusen BD, Spytek KA, Majumder K, Tchernev VT;
PI Padigaru M, Patturajan M, Burgess CE, Gangolli EA, Smithson G;
PI Rastelli L, Macdougall JR, Taupier RJ, Grosse WM, Szekeres ES;
PI Alsobrook JP;
XX
DR WPI; 2002-049278/06.
DR P-PSDB; AAM47673.
XX
PT Novel G-protein coupled receptor-related polypeptides and
PT polynucleotides for diagnosing, preventing and treating cardiomyopathy,
PT atherosclerosis, disorders related to cell signal processing and for
PT identifying modulators -
XX
PS Claim 8; Page 78; 227pp; English.
XX
CC The present invention relates to novel G-coupled protein-receptor related
CC proteins and coding sequences (MOLX, where X is a number from 1 to 10,
CC ABA04589-ABA04603 and AAM47659-AAM47673). MOLX proteins and coding

CC sequences are useful for treating or preventing a MOLX-associated
CC disorder, such as cardiomyopathy, atherosclerosis, disorders related to
CC cell signal processing and metabolic pathway modulation, diabetes and
CC cancer. Additionally, MOLX proteins and coding sequences are useful for
CC preventing and treating a variety of disorders including metabolic
CC disorders, nutritional oedema, chronic and hereditary pancreatitis,
CC obesity, infectious disease, anorexia, neurodegenerative disorders,
CC Alzheimer's disease, Parkinson's disease, stroke, immune disorders,
CC haematopoietic disorders and various dyslipidaemias, metabolic syndrome X
CC and wasting disorders associated with chronic diseases and cancers,
CC cardiac disorders, hypertension, hypercalcaemia, cirrhosis, angiogenesis
CC and wound healing, trauma, glomerulonephritis, hyper and hypothyroidism,
CC multiple sclerosis, lung diseases including asthma, Crohn's disease,
CC scleroderma, autoimmune diseases, developmental disorders and neural tube
CC defects. The present sequence is the coding sequence for human MOL10b,
CC which maps to chromosome 11. MOL10b protein is a cyclic nucleotide-gated
CC olfactory channel-like protein.
XX
SQ Sequence 2551 BP; 646 A; 668 C; 707 G; 530 T; 0 other;

(See Table 10E, page 79)

102(a)

alignment_scores:
Quality: 2984.00 Length: 575
Ratio: 5.190 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.826

alignment_block:
US-09-927-267-1 x ABA04603 ..

Align seg 1/1 to: ABA04603 from: 1 to: 2551

```
      1 MetSerGlnAspThrLysValLysThrThrGluSerSerProProAlaPr 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    779 ATGAGCCAGGACACCAAAGTGAAGACAACAGAGTCCAGTCCCCCAGCCCC 828

     17 oSerLysAlaArgLysLeuLeuProValLeuAspProSerGlyAspTyrT 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    829 ATCCAAGGCCAGGAAGTTGCTGCCTGTCCTGGACCCATCTGGGGATTACT 878

     34 yrTyrTrpTrpLeuAsnThrMetValPheProValMetTyrAsnLeuIle 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    879 ACTACTGGTGGCTGAACACAATGGTCTTCCCAGTCATGTATAACCTCATC 928

     51 IleLeuValCysArgAlaCysPheProAspLeuGlnHisGlyTyrLeuVa 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    929 ATCCTCGTGTGCAGAGCCTGCTTCCCCGACTTGCAGCACGGTTATCTGGT 978

     67 lAlaTrpLeuValLeuAspTyrThrSerAspLeuLeuTyrLeuLeuAspM 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    979 GGCCTGGTTGGTGCTGGACTACACGAGTGACCTGCTATACCTACTAGACA 1028

     84 etValValArgPheHisThrGlyPheLeuGluGlnGlyIleLeuValVal 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1029 TGGTGGTGCGCTTCCACACAGGATTCTTGGAACAGGGCATCCTGGTGGTG 1078

    101 AspLysGlyArgIleSerSerArgTyrValArgThrTrpSerPhePheLe 117
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1079 GACAAGGGTAGGATCTCGAGTCGCTACGTTCGCACCTGGAGTTTCTTCTT 1128

    117 uAspLeuAlaSerLeuMetProThrAspValValTyrValArgLeuGlyP 134
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1129 GGACCTGGCTTCCCTGATGCCCACAGATGTGGTCTACGTGCGGCTGGGCC 1178

    134 roHisThrProThrLeuArgLeuAsnArgPheLeuArgAlaProArgLeu 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1179 CGCACACACCCACCCTGAGGCTGAACCGCTTTCTCCGCGCGCCCCGCCTC 1228
                              157
    151 PheGluAlaPheAspArgThrGluThrArgThrAlaTyrProAsnAlaPh 167
        ||||||||||||||||||:::||||||||||||||||||||||||||||
   1229 TTCGAGGCCTTCGACCGCGCAGAGACCCGCACAGCTTACCCAAATGCCTT 1278

    167 eArgIleAlaLysLeuMetLeuTyrIlePheValValIleHisTrpAsnS 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
   1279 TCGCATTGCCAAGCTGATGCTTTACATTTTTGTCGTCATCCATTGGAACA 1328

    184 erCysLeuTyrPheAlaLeuSerArgTyrLeuGlyPheGlyArgAspAla 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1329 GCTGCCTATACTTTGCCCTATCCCGGTACCTGGGCTTCGGGCGTGACGCA 1378

    201 TrpValTyrProAspProAlaGlnProGlyPheGluArgLeuArgArgGl 217
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1379 TGGGTGTACCCGGACCCCGCGCAGCCTGGCTTTGAGCGCCTGCGGCGCCA 1428

    217 nTyrLeuTyrSerPheTyrPheSerThrLeuIleLeuThrThrValGlyA 234
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1429 GTACCTCTATAGCTTTTACTTCTCCACGCTGATACTGACTACAGTGGGCG 1478

    234 spThrProProProAlaArgGluGluGluTyrLeuPheMetValGlyAsp 250
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1479 ATACACCGCCGCCAGCCAGGGAAGAAGAGTACCTCTTCATGGTGGGCGAC 1528

    251 PheLeuLeuAlaValMetGlyPheAlaThrIleMetGlySerMetSerSe 267
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1529 TTCCTGCTGGCCGTCATGGGTTTCGCCACCATCATGGGTAGCATGAGCTC 1578

    267 rValIleTyrAsnMetAsnThrAlaAspAlaAlaPheTyrProAspHisA 284
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1579 TGTCATCTACAACATGAACACTGCAGATGCGGCTTTCTACCCAGATCATG 1628

    284 laLeuValLysLysTyrMetLysLeuGlnHisValAsnArgLysLeuGlu 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1629 CACTGGTGAAGAAGTACATGAAGCTGCAGCACGTCAACCGCAAGCTGGAG 1678

    301 ArgArgValIleAspTrpTyrGlnHisLeuGlnIleAsnLysLysMetTh 317
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1679 CGGCGAGTTATTGACTGGTATCAGCACCTGCAGATCAACAAGAAGATGAC 1728

    317 rAsnGluValAlaIleLeuGlnHisLeuProGluArgLeuArgAlaGluV 334
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1729 CAACGAGGTAGCCATCTTACAGCACTTGCCTGAGCGGCTGCGGGCAGAAG 1778

    334 alAlaValSerValHisLeuSerThrLeuSerArgValGlnIlePheGln 350
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1779 TGGCTGTGTCTGTGCACCTGTCCACTCTGAGCCGGGTGCAGATCTTTCAG 1828

    351 AsnCysGluAlaSerLeuLeuGluGluLeuValLeuLysLeuGlnProGl 367
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1829 AACTGTGAGGCCAGCCTGCTGGAGGAGCTGGTGCTGAAGCTGCAGCCCCA 1878

    367 nThrTyrSerProGlyGluTyrValCysArgLysGlyAspIleGlyGlnG 384
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1879 GACCTACTCACCAGGTGAATATGTATGCCGCAAAGGAGACATTGGCCAAG 1928

    384 luMetTyrIleIleArgGluGlyGlnLeuAlaValValAlaAspAspGly 400
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1929 AGATGTACATCATCCGAGAGGGTCAACTGGCCGTGGTGGCAGATGATGGT 1978

    401 IleThrGlnTyrAlaValLeuGlyAlaGlyLeuTyrPheGlyGluIleSe 417
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1979 ATCACACAGTATGCTGTGCTCGGTGCAGGGCTCTACTTTGGGGAGATCAG 2028

    417 rIleIleAsnIleLysGlyAsnMetSerGlyAsnArgArgThrAlaAsnI 434
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2029 CATCATCAACATCAAAGGGAACATGTCTGGGAACCGCCGCACAGCCAACA 2078

    434 leLysSerLeuGlyTyrSerAspLeuPheCysLeuSerLysGluAspLeu 450
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2079 TCAAGAGCCTAGGTTATTCAGACCTATTCTGCCTGAGCAAGGAGGACCTG 2128

    451 ArgGluValLeuSerGluTyrProGlnAlaGlnThrIleMetGluGluLy 467
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2129 CGGGAGGTGCTGAGCGAGTATCCACAAGCACAGACCATCATGGAGGAGAA 2178

    467 sGlyArgGluIleLeuLeuLysMetAsnLysLeuAspValAsnAlaGluA 484
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2179 AGGACGTGAGATCCTGCTGAAAATGAACAAGTTGGACGTGAATGCTGAGG 2228
```

```
     484 laAlaGluIleAlaLeuGlnGluAlaThrGluSerArgLeuArgGlyLeu 500
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2229 CAGCTGAGATCGCCCTGCAGGAGGCCACAGAGTCCCGGCTACGAGGCCTA 2278

     501 AspGlnGlnLeuAspAspLeuGlnThrLysPheAlaArgLeuLeuAlaGl 517
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2279 GACCAGCAGCTGGATGATCTACAGACCAAGTTTGCTCGCCTCCTGGCTGA 2328

     517 uLeuGluSerSerAlaLeuLysIleAlaTyrArgIleGluArgLeuGluT 534
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2329 GCTGGAGTCCAGCGCACTTAAGATTGCTTACCGCATTGAACGGCTGGAGT 2378

     534 rpGlnThrArgGluTrpProMetProGluAspLeuAlaGluAlaAspAsp 550
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2379 GGCAGACTCGAGAGTGGCCAATGCCCGAGGACCTGGCTGAGGCTGATGAC 2428

     551 GluGlyGluProGluGluGlyThrSerLysAspGluGluGlyArgAlaSe 567
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2429 GAGGGTGAGCCTGAGGAGGGAACTTCCAAAGATGAAGAGGGCAGGGCCAG 2478

     567 rGlnGluGlyProProGlyProGlu 575
         |||||||||||||||||||||||||
    2479 CCAGGAGGGACCCCCAGGTCCAGAG 2503

seq_name: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:ABA04602

seq_documentation_block:
ID   ABA04602 standard; cDNA; 1835 BP.
XX
AC   ABA04602;
XX
DT   21-FEB-2002  (first entry)
XX
DE   MOL10a coding sequence.
XX
KW   MOL; G-coupled protein-receptor; cardiomyopathy; atherosclerosis;
KW   cell signal processing; metabolic disorder; diabetes; cancer;
KW   neurodegenerative disorder; immune disorder; cardiac disorder;
KW   lung disease; autoimmune disease; developmental disorder; antidiabetic;
KW   Cytostatic; Neuroprotective; Antiatherosclerotic; Immunosuppressive;
KW   Gene therapy; Vaccine; antiinflammatory; MOL10a; ss.
XX
OS   Unidentified.
XX
FH   Key             Location/Qualifiers
FT   CDS             54..1790
FT                   /*tag=  a
FT                   /product= "MOL 10a"
XX
PN   WO200181578-A2.
XX
PD   01-NOV-2001.
XX
PF   26-APR-2001; 2001WO-US13578.
XX
PR   26-APR-2000; 2000US-200158P.
PR   28-APR-2000; 2000US-200613P.
PR   28-APR-2000; 2000US-200780P.
PR   01-MAY-2000; 2000US-201006P.
PR   01-MAY-2000; 2000US-201007P.
PR   01-MAY-2000; 2000US-201236P.
PR   01-MAY-2000; 2000US-201238P.
PR   02-MAY-2000; 2000US-201186P.
PR   03-MAY-2000; 2000US-201474P.
PR   03-MAY-2000; 2000US-201508P.
PR   25-JUL-2000; 2000US-220591P.
PR   15-SEP-2000; 2000US-232678P.
PR   22-JAN-2001; 2001US-263217P.
PR   30-JAN-2001; 2001US-265160P.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Vernet CAM,  Fernandes ER,  Gerlach V,  Shimkets RA,  Malyankar UM;
PI   Boldog FL,  Zerhusen BD,  Spytek KA,  Majumder K,  Tchernev VT;
PI   Padigaru M,  Patturajan M,  Burgess CE,  Gangolli EA,  Smithson G;
PI   Rastelli L,  Macdougall JR,  Taupier RJ,  Grosse WM,  Szekeres ES;
PI   Alsobrook JP;
XX
DR   WPI; 2002-049278/06.
DR   P-PSDB; AAM47672.
XX
PT   Novel G-protein coupled receptor-related polypeptides and
PT   polynucleotides for diagnosing, preventing and treating cardiomyopathy,
PT   atherosclerosis, disorders related to cell signal processing and for
PT   identifying modulators  -
XX
PS   Claim 8; Page 76; 227pp; English.
XX
CC   The present invention relates to novel G-coupled protein-receptor related
CC   proteins and coding sequences (MOLX, where X is a number from 1 to 10,
CC   ABA04589-ABA04603 and AAM47659-AAM47673). MOLX proteins and coding
CC   sequences are useful for treating or preventing a MOLX-associated
CC   disorder, such as cardiomyopathy, atherosclerosis, disorders related to
CC   cell signal processing and metabolic pathway modulation, diabetes and
CC   cancer. Additionally, MOLX proteins and coding sequences are useful for
CC   preventing and treating a variety of disorders including metabolic
CC   disorders, nutritional oedema, chronic and hereditary pancreatitis,
CC   obesity, infectious disease, anorexia, neurodegenerative disorders,
CC   Alzheimer's disease, Parkinson's disease, stroke, immune disorders,
CC   haematopoietic disorders and various dyslipidaemias, metabolic syndrome X
CC   and wasting disorders associated with chronic diseases and cancers,
CC   cardiac disorders, hypertension, hypercalcaemia, cirrhosis, angiogenesis
CC   and wound healing, trauma, glomerulonephritis, hyper and hypothyroidism,
CC   multiple sclerosis, lung diseases including asthma, Crohn's disease,
CC   scleroderma, autoimmune diseases, developmental disorders and neural tube
CC   defects. The present sequence is the coding sequence for MOL10a.
CC   MOL10a protein is a cyclic nucleotide-gated olfactory channel-like
CC   protein.
XX
SQ   Sequence 1835 BP; 420 A; 521 C; 521 G; 373 T; 0 other;


alignment_scores:
           Quality: 2957.50              Length:    578
             Ratio:  5.143                Gaps:      3
 Percent Similarity:  99.481   Percent Identity: 99.481

alignment_block:
 US-09-927-267-1 x ABA04602   ..

 Align seg 1/1  to: ABA04602  from: 1  to: 1835

       1 MetSerGlnAspThrLysValLysThrThrGluSerSerProProAlaPr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      54 ATGAGCCAGGACACCAAAGTGAAGACAACAGAGTCCAGTCCCCCAGCCCC 103

      17 oSerLysAla...ArgLysLeuLeuProValLeuAspProSerGlyAspT 33
         ||||||||||   ||||||||||||||||||||||||||||||||||||
     104 ATCCAAGGCCAGGAGGAAGTTGCTGCCTGTCCTGGACCCATCTGGGGATT 153

      33 yrTyrTyrTrpTrpLeuAsnThrMetValPheProValMetTyrAsnLeu 49
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     154 ACTACTACTGGTGGCTGAACACAATGGTCTTCCCAGTCATGTATAACCTC 203

      50 IleIleLeuValCysArgAlaCysPheProAspLeuGlnHisGlyTyrLe 66
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     204 ATCATCCTCGTGTGCAGGGCCTGCTTCCCCGACTTGCAGCACGGTTATCT 253

      66 uValAlaTrpLeuValLeuAspTyrThrSerAspLeuLeuTyrLeuLeuA 83
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     254 GGTGGCCTGGTTGGTGCTGGACTACACGAGTGACCTGCTATACCTACTAG 303

      83 spMetValValArgPheHisThr...GlyPheLeuGluGlnGlyIleLeu 98
         ||||||||||||||||||||||   |||||||||||||||||||||||||
     304 ACATGGTGGTGCGCTTCCACACAGGTGGATTCTTGGAACAGGGCATCCTG 353
```

```
  99 ValValAspLysGlyArgIleSerSerArgTyrValArgThrTrpSerPh 115
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 354 GTGGTGGACAAGGGTAGGATCTCGAGTCGCTACGTTCGCACCTGGAGTTT 403

 115 ePheLeuAspLeuAlaSerLeuMetProThrAspValValTyrValArgL 132
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 404 CTTCTTGGACCTGGCTTCCCTGATGCCCACAGATGTGGTCTACGTGCGGC 453

 132 euGlyProHisThrProThrLeuArgLeuAsnArgPheLeuArgAlaPro 148
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 454 TGGGCCCGCACACACCCACCCTGAGGCTGAACCGCTTTCTCCGCGCGCCC 503

 149 ArgLeuPheGluAlaPheAspArgThrGluThrArgThrAlaTyrProAs 165
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 504 CGCCTCTTCGAGGCCTTCGACCGCACAGAGACCCGCACAGCTTACCCAAA 553

 165 nAlaPheArgIleAlaLysLeuMetLeuTyrIlePheValValIleHisT 182
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 554 TGCCTTTCGCATTGCCAAGCTGATGCTTTACATTTTTGTCGTCATCCATT 603

 182 rpAsnSerCysLeuTyrPheAlaLeuSerArgTyrLeuGlyPheGlyArg 198
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 604 GGAACAGCTGCCTATACTTTGCCCTATCCCGGTACCTGGGCTTCGGGCGT 653

 199 AspAlaTrpValTyrProAspProAlaGlnProGlyPheGluArgLeuAr 215
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 654 GACGCATGGGTGTACCCGGACCCCGCGCAGCCTGGCTTTGAGCGCCTGCG 703

 215 gArgGlnTyrLeuTyrSerPheTyrPheSerThrLeuIleLeuThrThrV 232
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 704 GCGCCAGTACCTCTATAGCTTTTACTTCTCCACGCTGATACTGACTACAG 753

 232 alGlyAspThrProProProAlaArgGluGluGluTyrLeuPheMetVal 248
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 754 TGGGCGATACACCGCCGCCAGCCAGGGAAGAAGAGTACCTCTTCATGGTG 803

 249 GlyAspPheLeuLeuAlaValMetGlyPheAlaThrIleMetGlySerMe 265
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 804 GGCGACTTCCTGCTGGCCGTCATGGGTTTCGCCACCATCATGGGTAGCAT 853

 265 tSerSerValIleTyrAsnMetAsnThrAlaAspAlaAlaPheTyrProA 282
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 854 GAGCTCTGTCATCTACAACATGAACACTGCAGATGCGGCTTTCTACCCAG 903

 282 spHisAlaLeuValLysLysTyrMetLysLeuGlnHisValAsnArgLys 298
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 904 ATCATGCACTGGTGAAGAAGTACATGAAGCTGCAGCACGTCAACCGCAAG 953

 299 LeuGluArgArgValIleAspTrpTyrGlnHisLeuGlnIleAsnLysLy 315
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 954 CTGGAGCGGCGAGTTATTGACTGGTATCAGCACCTGCAGATCAACAAGAA 1003

 315 sMetThrAsnGluValAlaIleLeuGlnHisLeuProGluArgLeuArgA 332
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1004 GATGACCAACGAGGTAGCCATCTTACAGCACTTGCCTGAGCGGCTGCGGG 1053

 332 laGluValAlaValSerValHisLeuSerThrLeuSerArgValGlnIle 348
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1054 CAGAAGTGGCTGTGTCTGTGCACCTGTCCACTCTGAGCCGGGTGCAGATC 1103

 349 PheGlnAsnCysGluAlaSerLeuLeuGluGluLeuValLeuLysLeuGl 365
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1104 TTTCAGAACTGTGAGGCCAGCCTGCTGGAGGAGCTGGTGCTGAAGCTGCA 1153

 365 nProGlnThrTyrSerProGlyGluTyrValCysArgLysGlyAspIleG 382
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1154 GCCCCAGACCTACTCACCAGGTGAATATGTATGCCGCAAAGGAGACATTG 1203

 382 lyGlnGluMetTyrIleIleArgGluGlyGlnLeuAlaValValAlaAsp 398
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1204 GCCAAGAGATGTACATCATCCGAGAGGGTCAACTGGCCGTGGTGGCAGAT 1253

 399 AspGlyIleThrGlnTyrAlaValLeuGlyAlaGlyLeuTyrPheGlyGl 415
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1254 GATGGTATCACACAGTATGCTGTGCTCGGTGCAGGGCTCTACTTTGGGGA 1303

 415 uIleSerIleIleAsnIleLys...GlyAsnMetSerGlyAsnArgArgT 431
     ||||||||||||||||||||||   ||||||||||||||||||||||||
1304 GATCAGCATCATCAACATCAAAGGTGGGAACATGTCTGGGAACCGCCGCA 1353

 431 hrAlaAsnIleLysSerLeuGlyTyrSerAspLeuPheCysLeuSerLys 447
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1354 CAGCCAACATCAAGAGCCTAGGTTATTCAGACCTATTCTGCCTGAGCAAG 1403

 448 GluAspLeuArgGluValLeuSerGluTyrProGlnAlaGlnThrIleMe 464
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1404 GAGGACCTGCGGGAGGTGCTGAGCGAGTATCCACAAGCACAGACCATCAT 1453

 464 tGluGluLysGlyArgGluIleLeuLeuLysMetAsnLysLeuAspValA 481
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1454 GGAGGAGAAAGGACGTGAGATCCTGCTGAAAATGAACAAGTTGGACGTGA 1503

 481 snAlaGluAlaAlaGluIleAlaLeuGlnGluAlaThrGluSerArgLeu 497
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1504 ATGCTGAGGCAGCTGAGATCGCCCTGCAGGAGGCCACAGAGTCCCGGCTA 1553

 498 ArgGlyLeuAspGlnGlnLeuAspAspLeuGlnThrLysPheAlaArgLe 514
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1554 CGAGGCCTAGACCAGCAGCTGGATGATCTACAGACCAAGTTTGCTCGCCT 1603

 514 uLeuAlaGluLeuGluSerSerAlaLeuLysIleAlaTyrArgIleGluA 531
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1604 CCTGGCTGAGCTGGAGTCCAGCGCACTTAAGATTGCTTACCGCATTGAAC 1653

 531 rgLeuGluTrpGlnThrArgGluTrpProMetProGluAspLeuAlaGlu 547
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1654 GGCTGGAGTGGCAGACTCGAGAGTGGCCAATGCCCGAGGACCTGGCTGAG 1703

 548 AlaAspAspGluGlyGluProGluGluGlyThrSerLysAspGluGluGl 564
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1704 GCTGATGACGAGGGTGAGCCTGAGGAGGGAACTTCCAAAGATGAAGAGGG 1753

 564 yArgAlaSerGlnGluGlyProProGlyProGlu 575
     ||||||||||||||||||||||||||||||||||
1754 CAGGGCCAGCCAGGAGGGACCCCCAGGTCCAGAG 1787
```

```
seq_name: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAD09558

seq_documentation_block:
ID   AAD09558 standard; cDNA; 1995 BP.
XX
AC   AAD09558;
XX
DT   10-SEP-2001  (first entry)
XX
DE   Human transporter and ion channel-7 (TRICH-7) cDNA.
XX
KW   Human; transporter and ion channel-7; TRICH-7; vaccine; cystic fibrosis;
KW   gene therapy; amyotrophic lateral sclerosis; amnesia; muscular dystrophy;
KW   hypertension; angina; neurological disorder; asthma; bipolar disorder;
KW   dementia; depression; Alzheimer's disease; epilepsy; mood; arrhythmia;
KW   Pick's disease; ischaemic cerebrovascular disease; AIDS; anxiety; stroke;
KW   Huntington's disease; Parkinson's disease; cerebral neoplasm; allergy;
KW   demyelinating disease; mental disorder; Schizophrenia; polymyositis;
KW   muscle disorder; cardiomyopathy; cataract; myocarditis; Grave's disease;
KW   dermatomyositis; diabetes mellitus; immunological disorder; psoriasis;
KW   rheumatoid arthritis; Sjogren's syndrome; systemic lupus erythematosus;
KW   sickle cell anaemia; Wilson's disease; infertility; Cushing's disease;
KW   scleroderma; pulmonary artery stenosis; nootropic; Addison's disease;
KW   malabsorption syndrome; hypercholesterolaemia; cancer; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1995
```

```
FT                   /*tag=  a
FT                   /product= "Human TRICH-7 protein"
XX
PN   WO200146258-A2.
XX
PD   28-JUN-2001.
XX
PF   22-DEC-2000; 2000WO-US35095.
XX
PR   23-DEC-1999;   99US-0172000.
PR   14-JAN-2000; 2000US-0176083.
PR   21-JAN-2000; 2000US-0177332.
PR   28-JAN-2000; 2000US-0178572.
PR   02-FEB-2000; 2000US-0179758.
PR   10-FEB-2000; 2000US-0181625.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Baughn MR,  Burford N,  Au-Young J,  Lu DAM,  Yang J,  Reddy R,  Lal P;
PI   Hillman JL,  Azimzai Y,  Yue H,  Nguyen DB,  Yao MG,  Gandhi AR;
PI   Tang YT,  Khan FA;
XX
DR   WPI; 2001-418042/44.
DR   P-PSDB; AAE04894.
XX
PT   Novel human transporter and ion channel proteins useful for treating
PT   and preventing transport, neurological, muscle and immunological
PT   disorders -
XX
PS   Claim 5; Page 145; 160pp; English.
XX
CC   The present sequence is transporter and ion channel-7 (TRICH-7) cDNA.
CC   TRICH is used as vaccine. TRICH is useful for treating a disease or
CC   condition associated with decreased expression of functional TRICH,
CC   such as transport disorder including amyotrophic lateral sclerosis,
CC   cystic fibrosis, Becker's muscular dystrophy, Charcot-Marie Tooth
CC   disease, Duchenne muscular dystrophy, angina and hypertension,
CC   neurological disorders including Alzheimer's disease, amnesia, bipolar
CC   disorder, dementia, depression, epilepsy, ischaemic cerebrovascular
CC   disease, stroke, cerebral neoplasms, Pick's disease, Huntington's
CC   disease and Parkinson's disease, demyelinating diseases, mental disorders
CC   including mood, anxiety, Schizophrenia and seasonal affective disorder,
CC   muscle disorder including cardiomyopathy, myocarditis, polymyositis,
CC   dermatomyositis, arrhythmias and asthma and immunological disorders
CC   including AIDS, adult respiratory distress syndrome (ARDS), allergies,
CC   anaemia, diabetes mellitus, rheumatoid arthritis, scleroderma, Sjogren's
CC   syndrome, systemic lupus erythematosus and other diseases including
CC   sickle cell anaemia, Wilson's disease, cataracts, infertility, pulmonary
CC   artery stenosis, Grave's disease, Cushing's disease, Addison's disease,
CC   glucose-galactose malabsorption syndrome, hypercholesterolaemia, cancers
CC   psoriasis and viral, bacterial, fungal, helminthic and protozoal
CC   infections. TRICH DNA is useful in gene therapy and in diagnostic
CC   purposes.
XX
SQ   Sequence 1995 BP; 511 A; 511 C; 532 G; 441 T; 0 other;


alignment_scores:
            Quality: 1535.00              Length:    540
              Ratio:  3.513                Gaps:      3
 Percent Similarity:  80.926   Percent Identity: 53.889

alignment_block:
 US-09-927-267-1 x AAD09558   ..

 Align seg 1/1  to: AAD09558  from: 1  to: 1995

      18 SerLysAlaArgLysLeuLeuProValLeuAspProSerGlyAspTyrTy 34
         :::|||   :::      |||   |||||||||||||:::||||||:::||
     370 ACCAAGAAGAAATTTGAACTATTTGTCTTGGACCCAGCTGGGGATTGGTA 419

      34 rTyrTrpTrpLeuAsnThrMetValPheProValMetTyrAsnLeuIleI 51
         ||||   ||||||   ::::::::::::||||||:::||||||      :

     420 CTACTGCTGGCTATTTGTCATTGCCATGCCCGTCCTTTACAACTGGTGCC 469

      51 leLeuValCysArgAlaCysPheProAspLeuGlnHisGlyTyrLeuVal 67
         ::||||||:::|||||||||||||   ||||||||   ||||||   :::
     470 TGCTGGTGGCCAGAGCCTGCTTCAGTGACCTACAGAAAGGCTACTACCTG 519

      68 AlaTrpLeuValLeuAspTyrThrSerAspLeuLeuTyrLeuLeuAspMe 84
         :::|||||||||||||||||:::||||||:::::|||:::   ||||::
     520 GTGTGGCTGGTGCTGGATTATGTCTCAGATGTGGTCTACATTGCGGACCT 569

      84 tValValArgPheHisThrGlyPheLeuGluGlnGlyIleLeuValValA 101
         :  :::|||::::::|||||||||||||||||||||:::||||||   |
     570 CTTCATCCGATTGCGCACAGGTTTCCTGGAGCAGGGGCTGCTGGTCAAAG 619

     101 spLysGlyArgIleSerSerArgTyrValArgThrTrpSerPhePheLeu 117
         ||      ::::::   ::::::|||:::::|||   :::|||   |||
     620 ATACCAAGAAACTGCGAGACAACTACATCCACACCCTGCAGTTCAAGCTG 669

     118 AspLeuAlaSerLeuMetProThrAspValValTyrValArgLeuGlyPr 134
         |||:::||||||:::::|||||||||:::::|||      :::
     670 GATGTGGCTTCCATCATCCCCACTGACCTGATCTATTTTGCTGTGGACAT 719

     134 oHisThrProThrLeuArgLeuAsnArgPheLeuArgAlaProArgLeuP 151
          |||:::|||   :::|||:::||||||:::|||:::      |||:::|
     720 CCACAGCCCTGAGGTGCGCTTCAACCGCCTGCTGCACTTTGCCCGCATGT 769

     151 heGluAlaPheAspArgThrGluThrArgThrAlaTyrProAsnAlaPhe 167
         |||||   ||||||||||||||||||||||   |||||||||   |||
     770 TTGAGTTCTTTGACCGGACAGAGACACGCACCAACTACCCTAACATCTTC 819

     168 ArgIleAlaLysLeuMetLeuTyrIlePheValValIleHisTrpAsnSe 184
         ||||||:::::|||:::||||||||||:::|||:::||||||||||::
     820 CGCATCAGCAACCTTGTCCTCTACATCTTGGTCATCATCCACTGGAATGC 869

     184 rCysLeuTyrPheAlaLeuSerArgTyrLeuGlyPheGlyArgAspAlaT 201
         :|||:::|||:::|||:::|||:::   :::|||||||||   |||:::|
     870 CTGCATCTATTATGCCATCTCCAAATCCATAGGCTTTGGGGTCGACACCT 919

     201 rpValTyrProAspProAlaGlnProGlyPheGluArgLeuArgArgGln 217
         ||||||||||||:::   ::::::|||   :::      |||   |||:::
     920 GGGTTTACCCAAACATCACTGACCCTGAGTATGGCTACCTGGCTAGGGAA 969

     218 TyrLeuTyrSerPheTyrPheSerThrLeuIleLeuThrThrValGlyAs 234
         |||:::|||   :::|||:::|||||||||   ||||||||||:::|||::
     970 TACATCTATTGCCTTTACTGGTCCACACTGACTCTCACTACCATTGGGGA 1019

     234 pThrProProProAlaArgGluGluGluTyrLeuPheMetValGlyAspP 251
         :||||||||||||:::::::::|||||||||||||||||:::::   ||||
    1020 GACACCACCCCCTGTAAAGGATGAGGAGTACCTATTTGTCATCTTTGACT 1069

     251 heLeuLeuAlaValMetGlyPheAlaThrIleMetGlySerMetSerSer 267
         |||||::::::|||:::   ||||||||||||:::|||:::::::::|||
    1070 TCCTGATTGGCGTCCTCATCTTTGCCACCATCGTGGGAAATGTGGGCTCC 1119

     268 ValIleTyrAsnMetAsnThrAlaAspAlaAlaPheTyrProAspHisAl 284
         :::|||   |||||||||::::::   |||   |||
    1120 ATGATCTCCAACATGAATGCCACCCGGGCAGAGTTCCAGGCTAAGATCGA 1169

     284 aLeuValLysLysTyrMetLysLeuGlnHisValAsnArgLysLeuGluA 301
             ||||||   ||||||:::::::::   |||::::::   :::|||
    1170 TGCCGTGAAACACTACATGCAGTTCCGAAAGGTCAGCAAGGGGATGGAAG 1219

     301 rgArgValIleAspTrpTyrGlnHisLeuGlnIleAsnLysLysMetThr 317
           :::||||||   |||::::::::::|||      |||||||||   :::
    1220 CCAAGGTCATTAGGTGGTTTGACTACTTGTGGACCAATAAGAAGACAGTG 1269

     318 AsnGluValAlaIleLeuGlnHisLeuProGluArgLeuArgAlaGluVa 334
         :::|||      ||||||:::::||||||   :::||||||||||::
    1270 GATGAGCGAGAAATTCTCAAGAATCTGCCAGCCAAGCTCAGGGCTGAGAT 1319

     334 lAlaValSerValHisLeuSerThrLeuSerArgValGlnIlePheGlnA 351
         :|||::::::|||||||||||||||||||:::::|||:::||||||::::
    1320 AGCCATCAATGTCCACTTGTCCACACTCAAGAAAGTGCGCATCTTCCATG 1369
```

```
 351 snCysGluAlaSerLeuLeuGluGluLeuValLeuLysLeuGlnProGln 367
     ::|||||||||:::|||||   |||||||||||||||||||:::||||||
1370 ATTGTGAGGCTGGCCTGCTGGTAGAGCTGGTACTGAAACTCCGTCCTCAG 1419

 368 ThrTyrSerProGlyGluTyrValCysArgLysGlyAspIleGlyGlnGl 384
     ::::::|||||||||:::|||:::|||||||||||||||||||||:::||
1420 GTCTTCAGTCCTGGGGATTACATTTGCCGCAAAGGGGACATCGGCAAGGA 1469

 384 uMetTyrIleIleArgGluGlyGlnLeuAlaValValAlaAspAspGlyI 401
     |||||||||||||:::||||||::|||||||||||||||||||||||:
1470 GATGTACATCATTAAGGAGGGCAAACTGGCAGTGGTGGCTGATGATGGTG 1519

 401 leThrGlnTyrAlaValLeuGlyAlaGlyLeuTyrPheGlyGluIleSer 417
     ::|||||||||||||:::|||:::||||||      ||||||||||||||
1520 TGACTCAGTATGCTCTGCTGTCGGCTGGAAGCTGCTTTGGCGAGATCAGT 1569

 418 IleIleAsnIleLysGlyAsnMetSerGlyAsnArgArgThrAlaAsnIl 434
     |||:::|||||||||||||:::      ||||||||||||||||||||||
1570 ATCCTTAACATTAAGGGCAGTAAAATGGGCAATCGACGCACAGCTAATAT 1619

 434 eLysSerLeuGlyTyrSerAspLeuPheCysLeuSerLysGluAspLeuA 451
     |:::||||||||||||||||||||||||||||||||||:::||||||
1620 CCGCAGCCTGGGCTACTCAGATCTCTTCTGCTTGTCCAAGGATGATCTTA 1669

 451 rgGluValLeuSerGluTyrProGlnAlaGlnThrIleMetGluGluLys 467
       |||:::::::::|||||||||:::|||:::   ::::::||||||:::
1670 TGGAAGCTGTGACTGAGTACCCTGATGCCAAGAAAGTCCTAGAAGAGAGG 1719

 468 GlyArgGluIleLeuLeuLysMetAsnLysLeuAspValAsn......Al 482
     ||||||||||||||:::|||   :::   ||||||   |||      ||
1720 GGTCGGGAGATCCTCATGAAGGAGGGACTGCTGGATGAGAACGAAGTGGC 1769

 482 aGluAlaAlaGluIleAlaLeuGlnGluAlaThrGluSerArgLeuArgG 499
     |   :::   |||:::   :::||||||           :::|||
1770 AACCAGCATGGAGGTCGACGTGCAGGAG.....AAGCTAGGGC 1807

 499 lyLeuAspGlnGlnLeuAspAspLeuGlnThrLysPheAlaArgLeuLeu 515
       |||:::   :::::::::   |||   |||:::|||:::||||||||||
1808 AGCTGGAGACCAACATGGAAACCTTGTACACTCGCTTTGGCCGCCTGCTG 1857

 516 AlaGluLeuGluSerSerAlaLeuLysIleAlaTyrArgIleGluArgLe 532
     ||||||      ::::::      |||:::      ||||||      ||
1858 GCTGAGTACACGGGGGCCCAGCAGAAGCTCAAGCAGCGCATCACAGTTCT 1907

 532 uGluTrpGlnThrArgGluTrpProMetProGluAspLeuAlaGluAlaA 549
     ||||   :::   ::::::            ||||||        :::|
1908 GGAAACCAAGATGAAACAG......AACAATGAAGATGACTACCTGTCTG 1951

 549 spAspGluGlyGluProGlu 555
     ||      ::::::||||||
1952 ATGGGATGAACAGCCCTGAG 1971

seq_name: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL13353

seq_documentation_block:
ID   ABL13353 standard; cDNA; 2909 BP.
XX
AC   ABL13353;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 34541.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB69250.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 34541; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2909 BP; 729 A; 726 C; 804 G; 650 T; 0 other;


alignment_scores:
           Quality: 1235.00            Length:    516
             Ratio:  3.111              Gaps:      2
 Percent Similarity:  76.938   Percent Identity: 47.093

alignment_block:
 US-09-927-267-1 x ABL13353   ..

 Align seg 1/1  to: ABL13353  from: 1  to: 2909

  27 LeuAspProSerGlyAspTyrTyrTyrTrpTrpLeuAsnThrMetValPh 43
     :::||||||:::   :::   :::|||   ||||||      :::   ::
 422 GTCGATCCCACCCTGCAGTCGCATTACAGATGGCTGGCCATCGTTTCACT 471

  43 eProValMetTyrAsnLeuIleIleLeuValCysArgAlaCysPheProA 60
     :   |||:::||||||:::|||:::::|||   ||||||   |||   :
 472 GGCAGTGCTTTACAATATCATCTTCGTGGTGGGACGTGCGGTCTTCTGGG 521

  60 spLeuGlnHisGlyTyrLeuValAlaTrpLeuValLeuAspTyrThrSer 76
     ::::::::   :::      :::   |||   :::|||||||||
 522 AGATCAATAAGAGTGCTCCCGCCTTTTGGTACACGCTGGACTACCTGTGC 571

  77 AspLeuLeuTyrLeuLeuAspMetValValArgPheHisThrGlyPheLe 93
     |||::::::||||||||||||   :::|||:::::::|||   ||||||||
 572 GACTTTATCTATCTGCTGGATACGCTCGTCCACATGCACGAGGGATTTTT 621

  93 uGluGlnGlyIleLeuValValAspLysGlyArgIleSerSerArgTyrV 110
     |:::||||||:::||||||   |||      |||:::      :::|||
 622 GGACCAAGGTCTCCTCGTGCGGGATGCCTTTCGGCTGCGCCGGCACTATT 671

 110 alArgThrTrpSerPhePheLeuAspLeuAlaSerLeuMetProThrAsp 126
       :::|||   ::::::::::||||||:::   |||::::::||||||||
 672 TCCACACCAAGGGCTGGTACCTCGACGTCCTGTCCATGCTGCCCACGGAC 721

 127 ValValTyrValArgLeuGlyProHisThrProThr..............  138
     ::::::|||:::         |||:::|||   :::
 722 TTGGCCTACATCTGGTGGCCGCCGGAGACCTGCTCCAGTCTGTATCTGCC 771

 139 .............LeuArgLeuAsnArgPheLeuArgAlaProArgLeuP 151
                  :::||||||||||||:::||||||      ||||||:
```

```
 772 GTGTCCCGTGATAGTGCGTCTGAATCGTCTGCTCAGGATCAATCGACTGT 821

 151 heGluAlaPheAspArgThrGluThrArgThrAlaTyrProAsnAlaPhe 167
     ::|||   ||||||||||||||||||   |||:::|||||||||||||||
 822 GGGAGTGGTTCGATCGAACGGAGACAGCGACGGGATATCCGAATGCCTTC 871

 168 ArgIleAlaLysLeuMetLeuTyrIlePheValValIleHisTrpAsnSe 184
     ||||||:::|||:::::|||   |||:::|||:::||||||||||||::
 872 CGCATCTGCAAGGTGGTGCTGGCCATTTTGGTGCTGATCCACTGGAATGC 921

 184 rCysLeuTyrPheAlaLeuSerArgTyrLeuGlyPheGlyArgAspAlaT 201
     :|||:::|||||||||:::|||      :::||||||:::   |||:::|
 922 GTGCATGTACTTTGCCATAAGCTACGAAATCGGCTTCAGTTCGGACTCGT 971

 201 rpValTyrProAspProAlaGlnProGlyPheGluArgLeuArgArgGln 217
     |||||||         :::      :::   :::   |||:::||||||
 972 GGGTGTATAATCTGAATGGAACGAGGAAC...AATACGTTGCAGCGGCAA 1018

 218 TyrLeuTyrSerPheTyrPheSerThrLeuIleLeuThrThrValGlyAs 234
     |||:::|||||||||||||:::|||||||||   ||||||||||:::|||::
1019 TATATCTATAGTTTCTACTGGTCGACGCTAACGCTGACCACCATTGGCGA 1068

 234 pThrProProProAlaArgGluGluGluTyrLeuPheMetValGlyAspP 251
     :||||||   |||   ::::::   |||||||||||||:::|||:::||||
1069 GACTCCAACGCCGGAGAACGATGTGGAGTATCTGTTCGTGGTGGCCGATT 1118

 251 heLeuLeuAlaValMetGlyPheAlaThrIleMetGlySerMetSerSer 267
     |||||   :::|||:::   ||||||||||||:::|||::::::::|||
1119 TCCTGGCTGGCGTCCTAATCTTCGCCACCATTGTGGGTAACATTGGCTCC 1168

 268 ValIleTyrAsnMetAsnThrAlaAspAlaAlaPheTyrProAspHisAl 284
     :::|||   ||||||||||:::|||   :::   |||
1169 ATGATCTCCAACATGAATGTGGCCCGCGTGGAGTTCCAGAATCGCATGGA 1218

 284 aLeuValLysLysTyrMetLysLeuGlnHisValAsnArgLysLeuGluA 301
         ||||||:::||||||   :::::::::|||:::::::::||||||
1219 CGGAGTCAAGCAGTACATGGCCTTCCGGCGAGTGGGTCACGAGCTGGAGG 1268

 301 rgArgValIleAspTrpTyrGlnHisLeuGlnIleAsnLysLysMetThr 317
       |||||||||   |||:::   :::          ::::::
1269 CCCGGGTGATCCGCTGGTTTGCCTACACGTGGTCGCAGAGTGGTGCGCTG 1318

 318 AsnGluValAlaIleLeuGlnHisLeuProGluArgLeuArgAlaGluVa 334
     :::|||      :::|||      ||||||::::::|||:::||||||::
1319 GACGAGGAGCGTGTGCTGGCCGCTCTGCCGGACAAGCTGAAGGCGGAGAT 1368

 334 lAlaValSerValHisLeuSerThrLeuSerArgValGlnIlePheGlnA 351
     :|||::::::||||||::::::||||||::::::|||:::||||||::::
1369 CGCTATCCAGGTGCACATGGATACGCTGAAGCAGGTGCGCATCTTCCACG 1418

 351 snCysGluAlaSerLeuLeuGluGluLeuValLeuLysLeuGlnProGln 367
     ::   |||   :::||||||||||   ||||||||||||||:::   |||
1419 ACACGGAACCGGGTCTCCTCGAGGCCCTGGTGCTGAAGCTCAAACTGCAG 1468

 368 ThrTyrSerProGlyGluTyrValCysArgLysGlyAspIleGlyGlnGl 384
     ::::::|||||||||:::|||:::||||||||||||||||:::|||:::||
1469 GTCTTCAGTCCGGGCGACTATATCTGTCGCAAGGGTGACGTGGGCAAGGA 1518

 384 uMetTyrIleIleArgGluGlyGlnLeuAlaValValAlaAspAspGlyI 401
     ||||||||||:::::::::|||:::|||:::||||||:::|||||||||
1519 GATGTACATCGTGAAGCGTGGCAAGTTGTCCGTCGTCGGTGATGATGGCA 1568

 401 leThrGlnTyrAlaValLeuGlyAlaGlyLeuTyrPheGlyGluIleSer 417
     |||||      |||:::||||||||||||      |||||||||:::|||
1569 TCACCGTGCTGGCCACCTTGGGCGCCGGTTCGGTGTTCGGCGAGGTTTCC 1618

 418 IleIleAsnIleLysGlyAsnMetSerGlyAsnArgArgThrAlaAsnIl 434
     ::::::::::|||   ||||||   :::||||||||||||||||||||::
1619 GTACTCGAGATTGCGGGCAATCGGACGGGCAATCGGCGCACCGCCAATGT 1668

 434 eLysSerLeuGlyTyrSerAspLeuPheCysLeuSerLysGluAspLeuA 451
     ::::|||||||||||||||||||||||||||||:::|||:::||||||
1669 GAGATCCCTGGGCTACTCGGATCTCTTCTGTCTGGCGAAACGGGATTTGT 1718
```

```
 451 rgGluValLeuSerGluTyrProGlnAlaGlnThrIleMetGluGluLys 467
       |||:::|||||||:::||||||:::|||::::::   :::   ::::::
1719 GGGAGACCCTGTCCGACTATCCGGAGGCCAGATCCACGCTCACTCAGCGA 1768

 468 GlyArgGluIleLeuLeuLysMetAsnLysLeuAspValAsnAlaGluAl 484
     |||   ::::::|||   |||   :::   ||||||   :::      ||
1769 GGATGCCAGCTGCTGCGCAAGGATGGCTTACTGGACGAGCAAATATTCGC 1818

 484 aAlaGluIleAlaLeuGlnGluAlaThrGluSerArgLeuArgGlyLeuA 501
     |   :::      ::::::::::::   |||:::   ::::::   |||:
1819 TGATTCACAAAGGGTACACGACAGCATCGAGGGCGGCATCGAAAAGCTGG 1868

 501 spGlnGlnLeuAspAspLeuGlnThrLysPheAlaArgLeuLeuAlaGlu 517
     ::   ::::::::::::|||:::   ::::::||||||||||||||||||
1869 AGCTCTCGGTGGAGAACCTCAACATGCGGCTGGCCCGCCTCCTGGCTGAG 1918

 518 LeuGluSerSerAlaLeuLysIleAlaTyrArgIleGluArgLeuGlu 533
           :::|||      ||||||      |||:::   :::||||||
1919 TACACGGCAAGCCAGGCCAAGATCAAGCAGCGCCTCGCGAAGCTGGAG 1966
```

```
seq_name: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH98838

seq_documentation_block:
ID   AAH98838 standard; cDNA; 995 BP.
XX
AC   AAH98838;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Rat EST-derived coding sequence SEQ ID NO: 695.
XX
KW   Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW   tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW   diagnostics; forensic test; gene mapping; genetic disorder;
KW   biodiversity; gene therapy; nutrition; ss.
XX
OS   Rattus norvegicus.
XX
PN   WO200154477-A2.
XX
PD   02-AUG-2001.
XX
PF   25-JAN-2001; 2001WO-US02687.
XX
PR   25-JAN-2000; 2000US-0491404.
PR   17-JUL-2000; 2000US-0617746.
PR   03-AUG-2000; 2000US-0631451.
PR   15-SEP-2000; 2000US-0663870.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Zhou P,  Qian XB,  Wang Z,  Chen R,  Asundi V;
PI   Cao Y,  Drmanac RA,  Zhang J,  Werhman T;
XX
DR   WPI; 2001-476164/51.
DR   P-PSDB; AAM24179.
XX
PT   Isolated polypeptide for treatment of diseases, diagnostics, raising
PT   antibodies and research use -
XX
PS   Claim 1; Page 634; 1275pp; English.
XX
CC   The present invention provides the protein and coding sequences of novel
CC   proteins from a variety of organisms, including human, dog, cat, horse,
CC   cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC   urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC   from the organism of interest. They can be used in diagnostics,
CC   forensics, gene mapping, identification of mutations, to assess
CC   biodiversity and for nutritional purposes. The present sequence is a cDNA
CC   of the invention.
XX
SQ   Sequence 995 BP; 187 A; 291 C; 269 G; 248 T; 0 other;
```

```
alignment_scores:
            Quality:  959.50              Length:    194
              Ratio:  5.159                Gaps:      2
 Percent Similarity:  95.876   Percent Identity: 93.299

alignment_block:
 US-09-927-267-1 x AAH98838   ..

 Align seg 1/1  to: AAH98838  from: 1  to: 995

      20 AlaArgLysLeuLeuProValLeuAspProSerGlyAspTyrTyrTyrTr 36
         :::|||||||||||||||||||||||||||||||||||||||||||||||||
     118 AGTAGGAAGTTGCTGCCTGTCCTGGACCCATCTGGGGATTACTACTACTG 167

      36 pTrpLeuAsnThrMetValPheProValMetTyrAsnLeuIleIleLeuV 53
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     168 GTGGCTGAACACAATGGTCTTCCCAGTCATGTATAACCTCATCATCCTCG 217

      53 alCysArgAlaCysPheProAspLeuGlnHisGlyTyrLeuValAlaTrp 69
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     218 TGTGCAGAGCCTGCTTCCCCGACTTGCAGCACGGTTATCTGGTGGCCTGG 267

      70 LeuValLeuAspTyrThrSerAspLeuLeuTyrLeuLeuAspMetValVa 86
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     268 TTGGTGCTGGACTACACGAGTGACCTGCTATACCTACTAGACATGGTGGT 317

      86 lArgPheHisThrGlyPheLeuGluGlnGlyIleLeuValValAspLysG 103
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     318 GCGCTTCCACACAGGATTCTTGGAACAGGGCATCCTGGTGGTGGACAAGG 367

     103 lyArgIleSerSerArgTyrValArgThrTrpSerPhePheLeuAspLeu 119
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     368 GTAGGATCTCGAGTCGCTACGTTCGCACCTGGAGTTTCTTCTTGGACCTG 417

     120 AlaSerLeuMetProThrAspValValTyrValArgLeuGlyProHisTh 136
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     418 GCTTCCCTGATGCCCACAGATGTGGTCTACGTGCGGCTGGGCCCGCACAC 467

     136 rProThrLeuArgLeuAsnArgPheLeuArgAlaProArgLeuPheGluA 153
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     468 ACCCACCCTGAGGCTGAACCGCTTTCTCCGCGCGCCCCGCCTCTTCGAGG 517

     153 laPheAspArgThrGluThrArgThrAlaTyrProAsnAlaPheArgIle 169
         |||||||||||||||||||||||||||||||||||||||||||||||||   |||
     518 CCTTCGACCGCACAGAGACCCGCACAGCTTACCCAAATGCCTTTTGCATT 567

     170 AlaLysLeuMetLeuTyrIlePheValValIleHisTrpAsnSerCysLe 186
         :::|||||||||||||||||||||      ||||||||||||:::|||||
     568 GGCAAGCTGATGCTTTACATTTTTGGCCGCATCCATTGGAACAACTGCCT 617

     186 uTyrPheAlaLeuSerArgTyrLeuGlyPheGlyArgAsp.AlaTrpVal 202
         |||||||:::|||||||||||||||||||||||||||:::    ||||||
     618 ATACTTTTCCCTATCCCGGTACCTGGGCTTTGGGCGTGAACCCATGGGTG 667

     203 TyrProAspPro...AlaGlnProGlyPhe 211
         ||||||   |||      ||||||||||||
     668 TACCCCGGACCCCGGCGCCAACCTGGGTTT 697

seq_name: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL18085

seq_documentation_block:
ID   ABL18085 standard; DNA; 2109 BP.
XX
AC   ABL18085;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 5728.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 5728; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2109 BP; 485 A; 589 C; 605 G; 430 T; 0 other;


alignment_scores:
            Quality:  947.00              Length:    438
              Ratio:  2.905                Gaps:      4
 Percent Similarity:  74.429   Percent Identity: 42.237

alignment_block:
 US-09-927-267-1 x ABL18085   ..

 Align seg 1/1  to: ABL18085  from: 1  to: 2109

      40 ThrMetValPheProValMetTyrAsnLeuIleIleLeuValCysArgAl 56
         :::|||:::|||      :::|||||||:::   :::::::::   |||
      10 TCCATGGCCTTC......CTGTACAACTTCTGGGTGATAATCTACCGCTT 53

      56 aCysPheProAspLeuGlnHisGlyTyrLeuValAlaTrpLeuValLeuA 73
          :::|||   ::::::::::::      ::::::   |||:::   ||||
      54 CGCCTTCCAGGAGATCAATCGGCGAACGATCGCCATTTGGTTCTGTCTGG 103

      73 spTyrThrSerAspLeuLeuTyrLeuLeuAspMetValValArgPheHis 89
         |||||   ||||||:::||||||||||:::|||::::::   :::|||:::
     104 ACTACCTGTCCGACTTCCTGTATCTGATCGATATACTGTTCCACTTTCGC 153

      90 ThrGlyPheLeuGluGlnGlyIleLeuValValAspLysGlyArgIleSe 106
         ||||||:::||||||:::|||:::|||   :::|||      ::::::
     154 ACCGGATACCTGGAGGACGGGGTGCTGCAGACGGACGCTTTGAAGCTGCG 203

     106 rSerArgTyrValArgThrTrpSerPhePheLeuAspLeuAlaSerLeuM 123
          :::::::|||:::::::::      |||::::::|||         |||:
     204 CACCCACTACATGAACTCGACGATCTTCTACATCGACTGCCTGTGCCTGC 253

     123 etProThrAspValValTyrValArgLeuGlyProHisThrProThrLeu 139
         ::|||   |||   :::|||:::   :::|||      :::      |||
     254 TGCCGCTGGACTTTCTCTATTTGTCCATCGGC...TTTAACTCGATCCTG 300
```

```
 140 ArgLeuAsnArgPheLeuArgAlaProArgLeuPheGluAlaPheAspAr 156
 301 CGCAGCTTCCGGCTGGTGAAGATCTACCGGTTCTGGGCCTTCATGGATCG 350

 156 gThrGluThrArgThrAlaTyrProAsnAlaPheArgIleAlaLysLeuM 173
 351 CACCGAGCGGCACACCAACTACCCGAACCTGTTCCGGAGCACGGCCCTCA 400

 173 etLeuTyrIlePheValValIleHisTrpAsnSerCysLeuTyrPheAla 189
 401 TCCACTACCTGCTTGTGATATTCCATTGGAACGGTTGTCTCTACCACATC 450

 190 LeuSerArgTyrLeuGlyPheGlyArgAspAlaTrpValTyrProAspPr 206
 451 ATACACAAGAACAATGGGTTCGGATCACGCAACTGGGTCTACCACGACTC 500

 206 oAlaGlnProGlyPheGluArgLeuArgArgGlnTyrLeuTyrSerPheT 223
 501 GGAGTCGGCGGAC.........GTGGTCAAGCAGTATCTGCAGAGCTACT 541

 223 yrPheSerThrLeuIleLeuThrThrValGlyAspThrProProProAla 239
 542 ACTGGTGCACCCTGGCCCTGACCACCATCGGGGATCTTCCCAAGCCGCGC 591

 240 ArgGluGluGluTyrLeuPheMetValGlyAspPheLeuLeuAlaValMe 256
 592 TCCAAGGGCGAGTACGTCTTTGTTATCCTGCAGCTGCTCTTTGGCCTGAT 641

 256 tGlyPheAlaThrIleMetGlySerMetSerSerValIleTyrAsnMetA 273
 642 GCTCTTCGCCACGGTACTGGGGCACGTGGCCAACATTGTGACGTCAGTGA 691

 273 snThrAlaAspAlaAlaPheTyrProAspHisAlaLeuValLysLysTyr 289
 692 GTGCAGCGCGCAAGGAGTTTCAAGCCAAGCTGGATGGCGTGAAGACGTAC 741

 290 MetLysLeuGlnHisValAsnArgLysLeuGluArgArgValIleAspTr 306
 742 ATGCGGATGCGACGTGTGCCGAATCATCTGCAGGTGAAGGTCATCAAATG 791

 306 pTyrGlnHisLeuGlnIleAsnLysLysMetThrAsnGluValAlaIleL 323
 792 GTTCGATTACCTGTGGCTTACGCAAAAATGCTCGGACGAGGAGCGCGCCG 841

 323 euGlnHisLeuProGluArgLeuArgAlaGluValAlaValSerValHis 339
 842 TGTCCTGTCTTCCTGATAAATTAAAGGCTGAAATAGCAATTAACGTCCAT 891

 340 LeuSerThrLeuSerArgValGlnIlePheGlnAsnCysGluAlaSerLe 356
 892 TTAGATACGCTCAAGCGGGTGGAGATTTTCCAAAACACAGAGGCCGGCTT 941

 356 uLeuGluGluLeuValLeuLysLeuGlnProGlnThrTyrSerProGlyG 373
 942 CCTATGCGAACTGGTGCTGCGCCTGAGGCCCGTGCTCTTCTCGCCCGGCG 991

 373 luTyrValCysArgLysGlyAspIleGlyGlnGluMetTyrIleIleArg 389
 992 ACTACATCTGCAGAAAGGGCGAGGTGGGCAAGGAGATGTACATTGTGAAC 1041

 390 GluGlyGlnLeuAlaValValAlaAspAspGlyIleThrGlnTyrAlaVa 406
1042 CGAGGACGATTGCAGGTGGTGGCCGACAATGGAAAGACGGTGATGGCCTC 1091

 406 lLeuGlyAlaGlyLeuTyrPheGlyGluIleSerIleIleAsnIleLysG 423
1092 CCTGAAGGCTGGTTCCTATTTTGGCGAGATTAGTATACTCAATATG.... 1137

 423 lyAsnMetSerGlyAsnArgArgThrAlaAsnIleLysSerLeuGlyTyr 439
1138 ..GGCACCGCAGGCAACCGACGCACAGCCAGCGTTCGCTCAGTGGGATAC 1185

 440 SerAspLeuPheCysLeuSerLysGluAspLeuArgGluValLeuSerGl 456
```

```
1186 AGCGACCTCTTCGTCCTGAGCAAGAAGGACATGTGGGACGTGCTGAAGGA 1235

 456 uTyrProGlnAlaGlnThrIleMetGluGluLysGlyArgGluIleLeuL 473
1236 GTATCCGGCGGCGCGTGTTCGTCTGGAGTCGATAGCCGTCAAGCGATTGG 1285

 473 euLysMetAsnLys 477
1286 AGAAATACAAGAAG 1299
```

seq_name: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL14931

seq_documentation_block:
ID ABL14931 standard; cDNA; 5133 BP.
XX
AC ABL14931;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 39275.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB70828.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 39275; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 5133 BP; 1201 A; 1449 C; 1501 G; 982 T; 0 other;

alignment_scores:
Quality: 867.50 Length: 592
Ratio: 2.230 Gaps: 13
Percent Similarity: 65.709 Percent Identity: 33.784

alignment_block:
US-09-927-267-1 x ABL14931 ..

Align seg 1/1 to: ABL14931 from: 1 to: 5133



seq_name: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL13352

```
seq_documentation_block:
ID   ABL13352 standard; cDNA; 8585 BP.
XX
AC   ABL13352;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 34538.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB69249.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 34538; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 8585 BP; 2404 A; 1971 C; 1922 G; 2288 T; 0 other;


alignment_scores:
            Quality:  863.50              Length:    921
              Ratio:  2.439                Gaps:      4
 Percent Similarity:  38.436   Percent Identity: 23.779

alignment_block:
 US-09-927-267-1 x ABL13352   ..

 Align seg 1/1  to: ABL13352  from: 1  to: 8585

      91 GlyPheLeuGluGlnGlyIleLeuValValAspLysGlyArgIleSerSe 107
         |||||||||:::||||||:::||||||   |||     |||:::
    3817 GGATTTTTGGACCAAGGTCTCCTCGTGCGGGATGCCTTTCGGCTGCGCCG 3866

     107 rArgTyrValArgThrTrpSerPheLeuAspLeuAlaSerLeuMetP 124
         :::|||   :::|||   ::::::::::|||||||:::   |||:::::::|
    3867 GCACTATTTCCACACCAAGGGCTGGTACCTCGACGTCCTGTCCATGCTGC 3916

     124 roThrAspValValTyrValArgLeuGlyProHisThrProThr..... 138
         ||||||||:::::::|||:::         |||:::|||   :::
    3917 CCACGGACTTGGCCTACATCTGGTGGCCGCCGGAGACCTGCTCCAGTCTG 3966

     139 .....LeuArgLeuAsnArgPheLeuArgAlaPr 148
                          :::||||||||||||||:::||||||
    3967 TATCTGCCGTGTCCCGTGATAGTGCGTCTGAATCGTCTGCTCAGGATCAA 4016

     148 oArgLeuPheGluAlaPheAspArgThrGluThrArgThrAlaTyrProA 165
          ||||||:::|||   ||||||||||||||||||   |||:::|||||||
    4017 TCGACTGTGGGAGTGGTTCGATCGAACGGAGACAGCGACGGGATATCCGA 4066

     165 snAlaPheArgIleAlaLysLeuMetLeuTyrIlePheValValIleHis 181
         |||||||||||||||:::|||:::::|||   |||:::|||:::||||||
    4067 ATGCCTTCCGCATCTGCAAGGTGGTGCTGGCCATTTTGGTGCTGATCCAC 4116

     182 TrpAsnSerCysLeuTyrPheAlaLeuSerArgTyrLeuGlyPheGlyAr 198
         ||||||:::|||:::|||||||||:::|||      :::|||||||:::
    4117 TGGAATGCGTGCATGTACTTTGCCATAAGCTACGAAATCGGCTTCAGTTC 4166

     198 gAspAlaTrpValTyrProAspProAlaGlnProGlyPheGluArgLeuA 215
          |||:::|||||||||          :::      :::   :::   |||:
    4167 GGACTCGTGGGTGTATAATCTGAATGGAACGAGGAAC...AATACGTTGC 4213

     215 rgArgGlnTyrLeuTyrSerPheTyrPheSerThrLeuIleLeuThrThr 231
         ::|||||||||:::|||||||||||||:::|||||||||   |||||||||
    4214 AGCGGCAATATATCTATAGTTTCTACTGGTCGACGCTAACGCTGACCACC 4263

     232 ValGlyAspThrProProProAlaArgGluGluGluTyrLeuPheMetVa 248
         :::|||:::||||||   |||   ::::::   ||||||||||||:::||
    4264 ATTGGCGAGACTCCAACGCCGGAGAACGATGTGGAGTATCTGTTCGTGGT 4313

     248 lGlyAspPheLeuLeuAlaValMetGlyPheAlaThrIleMet..... 262
         |:::|||||||||   :::|||:::   ||||||||||||:::
    4314 GGCCGATTTCCTGGCTGGCGTCCTAATCTTCGCCACCATTGTGGGTGAGT 4363

     262 ..................................................... 262

    4364 TGTGAAAGGAATGGGTCATTTAACTGGGTTGATAACCTAGAAAATCCCTT 4413

     262 ..................................................... 262

    4414 TCGGGCTCCCTCAGTCCCTCAGTTTGCTCGTGTTTACGCAAATATTTCGC 4463

     262 ..................................................... 262

    4464 CTGATCGTTTAATTGGTTCCAGAGTGACATCTGTGACATCAATGGGTATT 4513

     262 ..................................................... 262

    4514 AATTTACAGCCACGCACGTCAAGGACCCACACGTACACCCCGACCTTCGC 4563

     262 ..................................................... 262

    4564 CGAACCACCGACCCACATTATTCATTACGCCTTCAGTTGTGTCCTCTTCG 4613

     262 ..................................................... 262

    4614 GTGGTTCCATGCCATTTTGTTTTGTTTTGCCGGGTTGATTTGTAGCCGTC 4663

     262 ..................................................... 262

    4664 ACCCCCCCATCCCCTCACATCATCATCTCCCACCGCTTCTGTTTTTTTTT 4713

     262 ..................................................... 262

    4714 TTCGTTTTTGGCAAACGTGCCTTTTTATCATTTTTGTCGTCGTCCTTATC 4763

     262 ..................................................... 262

    4764 GCCGCCATTACCGCCGCCCAAATGGGTTGGCGGTGTCAAGGAGTAAGGCA 4813

     262 ..................................................... 262

    4814 CATTCCTGCATTCCCAACAGCTGCCAGACTCCCCAGATCCCCATCATTTG 4863
```

```
 262 ..................................................  262

4864 ACTGCTCCATCCCCACGACCTGGCTCTGCAAACAAATTTACTTACTGTGA 4913

 262 ..................................................  262

4914 TGAATGCCGAGTTCGTTTGGACAGGACAATTTGACGGGCACACTCGAAGA 4963

 262 ..................................................  262

4964 AATAGATTAATAGTACTAGGAAATTAGGCAGGAAATCTCAGAAGAGAGAG 5013

 262 ..................................................  262

5014 ACTAGATTGCATGTCCTTACATATTTACAAAAGTATGAGCAGATATGTGA 5063

 262 ..................................................  262

5064 ATTACATAGTGTTGTTAGTTAAATCAAGTTGGTTCTATTAGTATAGTATT 5113

 262 ..................................................  262

5114 AAATCTAATTAAAATTAAATCAGCTTTCTTAAAACAATAAGGCTGTTCAT 5163

 262 ..................................................  262

5164 AGCCTATAATAGCTCCTTATGTTACTCATTATTAATTTTTCCTGCCAAAA 5213

 262 ..................................................  262

5214 ATTTATGCAAGTACAAAATATGACTCTATTCATTAGGGCCTTAAAGTCAA 5263

 262 ..................................................  262

5264 CCAAGTTAGATTTCAAAATATCCTGAGCTAAGTATTTGTTTTTCGCTGAG 5313

 262 ..................................................  262

5314 TGCCCCTAGGCATCTTCTCCATTTCGTGGAATCGAGCATGTGTGTGTATT 5363

 262 ..................................................  262

5364 TGACTCCGGCACCGGAAGTGCCGCCATTAGTCGCCTTTGTGCCACGTTAT 5413

 262 ..................................................  262

5414 CCGCACGCCGGAATCCATGAGTGAAGCTTCGGCCAGCTTCGGTTTCACTG 5463

 262 ..................................................  262

5464 CATACCATGTGGGCGGGCGAACCGGTTATGAAATGGCCAAGATGAATTGC 5513

 262 ..................................................  262

5514 TGCCACTCTGCTTGTTCATTAACCAAGTTTCCCTAACACACTGCGCCCAT 5563

 263 .....GlySerMetSerSerValIleTyrAsnMetAsnThrAlaAspAla 277
          |||:::::::::|||:::|||   |||||||||:::|||   :::
5564 TTCCAGGTAACATTGGCTCCATGATCTCCAACATGAATGTGGCCCGCGTG 5613

 278 AlaPheTyrProAspHisAlaLeuValLysLysTyrMetLysLeuGlnHi 294
       |||                |||||:::|||||   ::::::::
5614 GAGTTCCAGAATCGCATGGACGGAGTCAAGCAGTACATGGCCTTCCGGCG 5663

 294 sValAsnArgLysLeuGluArgArgValIleAspTrpTyrGlnHisLeuG 311
     :|||::::::::::|||||   |||||||||   |||:::   :::
5664 AGTGGGTCACGAGCTGGAGGCCCGGGTGATCCGCTGGTTTGCCTACACGT 5713

 311 lnIleAsnLysLysMetThrAsnGluValAlaIleLeuGlnHisLeuPro 327
       ::::::          :::|||      :::|||      ||||||
5714 GGTCGCAGAGTGGTGCGCTGGACGAGGAGCGTGTGCTGGCCGCTCTGCCG 5763

 328 GluArgLeuArgAlaGluValAlaValSerValHisLeuSerThrLeuSe 344
```

```
      ::::::|||:::|||||||:::|||:::::::|||||||:::::::|||||||::
5764 GACAAGCTGAAGGCGGAGATCGCTATCCAGGTGCACATGGATACGCTGAA 5813

 344 rArgValGlnIlePheGlnAsnCysGluAlaSerLeuLeuGluGluLeuV 361
     ::::|||:::||||||:::::   |||   :::||||||||||   ||||
5814 GCAGGTGCGCATCTTCCACGACACGGAACCGGGTCTCCTCGAGGCCCTGG 5863

 361 alLeuLysLeuGlnProGlnThrTyrSerProGlyGluTyrValCysArg 377
     ||||||||||:::   |||::::::|||||||||:::|||:::||||||
5864 TGCTGAAGCTCAAACTGCAGGTCTTCAGTCCGGGCGACTATATCTGTCGC 5913

 378 LysGlyAspIleGlyGlnGluMetTyrIleIleArgGluGlyGlnLeuAl 394
     |||||||||:::|||:::||||||||||||:::::::::|||:::|||::
5914 AAGGGTGACGTGGGCAAGGAGATGTACATCGTGAAGCGTGGCAAGTTGTC 5963

 394 aValValAlaAspAspGlyIleThrGlnTyrAlaValLeuGlyAlaGlyL 411
     :||||||:::||||||||||||||      |||:::||||||||||||
5964 CGTCGTCGGTGATGATGGCATCACCGTGCTGGCCACCTTGGGCGCCGGTT 6013

 411 euTyrPheGlyGluIleSerIleIleAsnIleLysGlyAsnMetSerGly 427
          |||||||||:::|||:::::::::|||   ||||||   :::|||
6014 CGGTGTTCGGCGAGGTTTCCGTACTCGAGATTGCGGGCAATCGGACGGGC 6063

 428 AsnArgArgThrAlaAsnIleLysSerLeuGlyTyrSerAspLeuPheCy 444
     ||||||||||||||||:::::|||||||||||||||||||||||||||||
6064 AATCGGCGCACCGCCAATGTGAGATCCCTGGGCTACTCGGATCTCTTCTG 6113

 444 sLeuSerLysGluAspLeuArgGluValLeuSerGluTyrProGlnAlaG 461
     ||||:::|||:::|||||||   |||:::|||||||:::|||||||:::|||:
6114 TCTGGCGAAACGGGATTTGTGGGAGACCCTGTCCGACTATCCGGAGGCCA 6163

 461 lnThrIleMetGluGluLysGlyArgGluIleLeuLeuLysMetAsnLys 477
     :::::   :::   ::::::|||   :::::::|||   |||   :::
6164 GATCCACGCTCACTCAGCGAGGATGCCAGCTGCTGCGCAAGGATGGCTTA 6213

 478 LeuAspValAsnAlaGluAlaAlaGluIle..................... 487
     ||||||   :::      |||:::::::::
6214 CTGGACGAGCAAATATTCGCTGGTAAGTTCGATTAGTACATTTAGGTATC 6263

 487 ..................................................  487

6264 AAATATTACGTATACGTCATGTTGTTCAAAACGTAAATGGAAGGTATTCC 6313

 487 ..................................................  487

6314 ACTATTTTCTCATTCGTCCTCATTATTTCTGTCTACAGTGCACTTCTAGA 6363

 487 ..................................................  487

6364 TGTTGCATTATATTATGTAACTGTTAGATATATTAAGATTAATATGAAAT 6413

 488 .......................AlaLeuGlnGluAlaThrGluSerA 496
                            ::::::::::::   |||:::
6414 TCAAACTCCATTTCCAGATTCACAAAGGGTACACGACAGCATCGAGGGCG 6463

 496 rgLeuArgGlyLeuAspGlnGlnLeuAspAspLeuGlnThrLysPheAla 512
       ::::::   |||:::   ::::::::::::|||:::   ::::::|||
6464 GCATCGAAAAGCTGGAGCTCTCGGTGGAGAACCTCAACATGCGGCTGGCC 6513

 513 ArgLeuLeuAlaGluLeuGluSerSerAlaLeuLysIleAlaTyrArgIl 529
     ||||||||||||||      :::|||      ||||||      |||::
6514 CGCCTCCTGGCTGAGTACACGGCAAGCCAGGCCAAGATCAAGCAGCGCCT 6563

 529 eGluArgLeuGlu 533
     :   :::|||||
6564 CGCGAAGCTGGAG 6576

seq_name: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL19457

seq_documentation_block:
ID   ABL19457 standard; DNA; 3275 BP.
XX
AC   ABL19457;
```

```
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 9844.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 9844; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3275 BP; 843 A; 849 C; 873 G; 710 T; 0 other;


alignment_scores:
           Quality:  781.50            Length:    618
             Ratio:  2.112               Gaps:     13
 Percent Similarity:  59.871   Percent Identity: 31.230

alignment_block:
 US-09-927-267-1 x ABL19457   ..

 Align seg 1/1  to: ABL19457  from: 1  to: 3275

      10 ThrGluSerSerProProAlaProSerLysAlaArgLysLeuLeuPro.. 25
         |||:::|||      |||:::|||:::        :::|||:::|||
    1380 ACCAAGAGTCTGCACCCTTCGCCGCAGCACAAGGTGCGTCTGGTGCCGGC 1429

      25 ..................................................  25

    1430 TGGTCAATCACCCCATCGTGGAAGACTCTTCGAAGCGGATACTCCAAGGT 1479

      25 ..................................................  25

    1480 CCAATGTGTGGCTCTGTTCCAGTTTGTGCGGCGCCAACAATGATGAGCGA 1529

      26 ValLeuAspProSerGlyAspTyrTyrTyrTrpTrpLeuAsnThrMetVa  42
         :::|||||||||:::|||     |||      ||||||   ::::::
    1530 ACCCTGGATCCCCAGGGCAAGATCTACATCTCCTGGCTATGCGTTGTATC 1579

      42 lPheProValMetTyrAsnLeuIleIleLeuValCysArgAlaCysPheP  59

         :::      :::||||||      ::::::      ||||||   ||||
    1580 GCTATCGTTTCTCTATAACGCCTGGGTAATCCCGCTGCGTGCCTCGTTTC 1629

      59 roAspLeuGlnHisGlyTyrLeuValAlaTrpLeuValLeuAspTyrThr 75
         ||                          ||||||:::   |||:::
    1630 CCTTTCAAACCAAGGAGAACACCAATATCTGGCTGGCCTGCGACTTTTGC 1679

      76 SerAspLeuLeuTyrLeuLeuAspMetValValArgPheHis...ThrGl 91
         :::|||:::::::||||||||||||:::|||         |||   :::
    1680 GCTGACATCATATATCTACTGGATGTGGTCTTCTTTAAACATCGGGTTAT 1729

      91 yPheLeuGluGlnGlyIleLeuValValAspLysGlyArgIleSerSerA 108
          :::|||   :::|||:::   |||   :::|||:::         ::::
    1730 GTACCTTTTTGAGGGATTCTGGGTGAAAAACAAGAACCTCACCAGAAAGA 1779

     108 rgTyrValArgThrTrpSerPhePheLeuAspLeuAlaSerLeuMetPro 124
         ::|||:::|||      :::|||   |||||||||   :::|||:::|||
    1780 ACTACATGCGAAAACTGCAGTTTAAGCTGGATCTGCTGGCCCTTCTTCCG 1829

     125 ThrAspValValTyrValArgLeuGlyProHisThrProThrLeuArgLe 141
            ::::::::::|||   :::||||||   ::::::       ||||||::
    1830 CTGGAGCTATTGTACTTTAAACTCGGAACACAGGCGGTTTGGTTGCGTTT 1879

     141 uAsnArgPheLeuArgAlaProArgLeuPheGluAlaPheAspArgThrG 158
         :  ||||||::::::         ::::::|||:::|||        :
    1880 TCCAAGGTTCTTTAAGATTCAGAGCTTCTGGGAAGTGTTTCGCCTTTTGG 1929

     158 luThrArgThrAlaTyrProAsnAlaPheArgIleAlaLysLeuMetLeu 174
         ::         :::   |||:::      |||:::||||||   :::
    1930 ATCGAGTCATATCTTCTCCGCACTTTGTTCGAGTGGCCAAAACCCTAACC 1979

     175 TyrIlePheValValIleHisTrpAsnSerCysLeuTyrPheAlaLeuSe 191
         |||::::::   :::||||||   :::::::::||||||:::|||   ||
    1980 TACATGCTCTACATGATCCACATCACGGCTGCTCTGTACTACGCCTATAG 2029

     191 rArgTyrLeuGlyPheGlyArgAspAlaTrpValTyrProAspProAlaG 208
         |   |||   |||:::|||::::::   ||||||:::        ::::
    2030 TGACTACCAGGGTCTGGGACAGAATCGTTGGGTCTTCAGCGGCAAAGGGC 2079

     208 lnProGlyPheGluArgLeuArgArgGlnTyrLeuTyrSerPheTyrPhe 224
         ::|||                       |||:::      |||   |||
    2080 ATCCC.....................TATGTAAGGTGCTTTGCTTTT 2105

     225 SerThrLeuIleLeuThrThrValGlyAspThrProProProAlaArgGl 241
         :::|||         |||:::::::|||   :::|||   |||   |||::
    2106 GCCACCAAGACGGCAACTTCGATAGGAAAAAACCCCAAGCCGGAACGTCA 2155

     241 uGluGluTyrLeuPheMetValGlyAspPheLeuLeuAlaValMetGlyP 258
         :  ||||||:::||||||:::      :::|||:::::::|||:::   |
    2156 GGGAGAGTACGTCTTCATGACGGTGGCCTGGCTAATGGGCGTCTTTGTGT 2205

     258 heAlaThrIleMetGlySerMetSerSerValIle......TyrAsnMet 272
         |||||   ::::::|||::::::   :::::::|||         :::
    2206 TCGCTCTGCTTATCGGTCAGATCAGGGATATCATCTCGACAGCCACTCGG 2255

     273 AsnThrAlaAspAlaAlaPheTyrProAspHisAlaLeuValLysLysTy 289
         |||       :::                |||:::::::|||      :::||
    2256 AATAAGCATGAGTATCGCCAGCTGGAGGATGAGACACTG......GAGTA 2299

     289 rMetLysLeuGlnHisValAsnArgLysLeuGluArgArgValIleAspT 306
         ||||:::     ::::::::::::::::::::::   ||||||      |
    2300 CATGCGACGCCTCAATCTCTCGCAGGAGGTGCAGTCGAGGGTCAAAATGT 2349

     306 rpTyrGlnHisLeuGlnIleAsnLysLysMetThrAsnGluValAlaIle 322
         ||:::|||           :::::::::      :::|||      |||
    2350 GGTTCCAGTTTACATGGGAGCAGCAGCGCACTTTGGATGAATCCAACATA 2399

     323 LeuGlnHisLeuProGluArgLeuArgAlaGluValAlaValSerValHi 339
         |||:::   ||||||   :::|||:::::::::::::|||:::|||||||
    2400 TTGGATGCTTTGCCCATCAATTTGAAAACCGATATCGCCATCTCCGTGCA 2449

     339 sLeuSerThrLeuSerArgValGlnIlePheGlnAsnCysGluAlaSerL 356
         |::::::|||||||||:::||||||:::|||   :::|||||||   :::|
```

```
 2450 CATCCAAACGCTCTCCAAGGTGCAACTCTTTGCTGATTGCGAGGAGGCTT 2499

  356 euLeuGluGluLeuValLeuLysLeuGlnProGlnThrTyrSerProGly 372
      |||||::::::|||||||||||||||||:::      |||:::   ||||||
 2500 TACTCAGAGATCTGGTTCTAAAACTAAGGGCAGTCACATTTTTGCCAGGA 2549

  373 GluTyrValCysArgLysGlyAspIleGlyGlnGluMetTyrIleIleAr 389
      ::::::|||||||||||||||||:::::::|||:::|||||||||||||:::::
 2550 GATTTCGTTTGCCGCAAAGGCGAAGTGGGCAGAGAGATGTATATCGTGAA 2599

  389 gGluGlyGlnLeuAlaValVal.........AlaAspAspGlyIleThrG 403
      :   ||||||:::   |||:::          :::::::|||   :::
 2600 ACTGGGACAGGTTCAAGTGATGGGTGGTCCTAGTAGCGATGTAGTG.... 2645

  403 lnTyrAlaValLeuGlyAlaGlyLeuTyrPheGlyGluIleSerIleIle 419
          |||:::|||      |||      |||||||||||||||:::::
 2646 ..CTGGCCACTCTAACCGAAGGTTCCGTGTTCGGGGAGATCAGTCTGCTG 2693

  420 AsnIleLysGlyAsnMetSerGlyAsnArgArgThrAlaAsnIleLysSe 436
      :::|||:::|||         ::::::||||||||||||:::::::::||
 2694 GGAATAAATGGA.........GCCGATCGCAGAACAGCAGATGTCCGATC 2734

  436 rLeuGlyTyrSerAspLeuPheCysLeuSerLysGluAspLeuArgGluV 453
      |   ||||||||||:::||||||   ||||||||||:::|||||||:::||||
 2735 CAAGGGCTACTCCAACCTGTTCGTGCTGTCCAAATCGGATTTAAACGAGG 2784

  453 alLeuSerGluTyrProGlnAlaGlnThrIleMetGluGluLysGlyArg 469
      ||::::::   ||||||   ||||||:::|||:::::::::::::::|||
 2785 TCATTGCCTACTATCCCACGGCCCAGGCGATTCTCAAAAAGCGTGCTCGT 2834

  470 GluIleLeuLeuLys......MetAsnLysLeuAspValAsnAlaGluAl 484
      :::::::::   |||       ::::::  :::   :::   |||
 2835 CAGCTAATGCGAAAAAATGCAGCCAGAGAGCGGGAAGAGGAGCGAGAGCG 2884

  484 aAlaGluIleAlaLeuGln.............................G 491
       |||:::   |||||||||
 2885 AGCACGCAGCGCCCTCCAGGCAGATGTGGTTATCGGCAATCCCAAAACAC 2934

  491 luAlaThrGluSerArgLeuArgGlyLeuAspGlnGlnLeuAspAspLeu 507
          |||      :::|||                                |||
 2935 CAGAGACGGCGCCCAAACTT.........................CTG 2957

  508 GlnThrLysPheAlaArgLeuLeuAlaGluLeuGluSerSerAlaLeuLy 524
      ||||||   :::      |||         :::|||||   |||:::
 2958 CAGACAGTTATCCAGGCGCTG......CCGTTTGAATCGCCCGCCGTGGT 3001

  524 sIleAlaTyrArgIleGluArgLeuGluTrpGlnThrArgGluTrpProM 541
       :::      |||   ::::::         :::   |||:::      :
 3002 GCTCATCACGAGAGGTTCGAAGAGGATGCGCCGCAAGCGCCAGTCCGTGC 3051

  541 etProGluAspLeuAlaGluAlaAspAspGluGlyGluProGluGluGly 557
      ::   |||   ::::::|||         |||               |||
 3052 AGATGGAGACCATTGTTGAGCCCAAGATGGAG.........GTGACAGGC 3092

  558 ThrSerLysAspGluGluGlyArgAlaSerGlnGluGlyProProGlyPr 574
      :::::::::   ::::::  ::::::   :::   |||      :::||
 3093 GCTGCCGAGACGCAAGATCAAAAGGGCTTAAAGACAGGACGATGTTCGCC 3142

  574 oGlu 575
      |:::
 3143 AGAT 3146

seq_name: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS91658

seq_documentation_block:
ID   AAS91658 standard; cDNA; 3811 BP.
XX
AC   AAS91658;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #27462.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG27471.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 27462; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3811 BP; 812 A; 1093 C; 1102 G; 804 T; 0 other;


alignment_scores:
            Quality:  714.00            Length:    563
              Ratio:  2.017               Gaps:     16
 Percent Similarity:  62.877   Percent Identity: 31.261

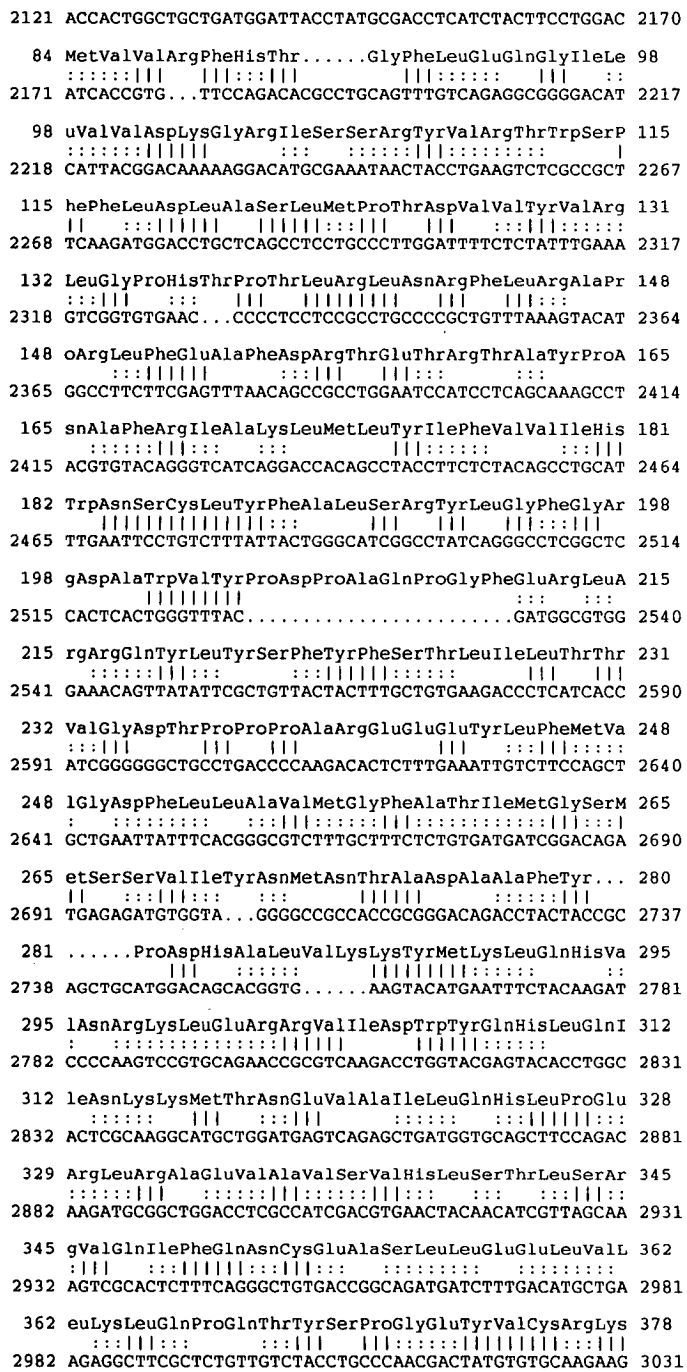
alignment_block:
 US-09-927-267-1 x AAS91658   ..

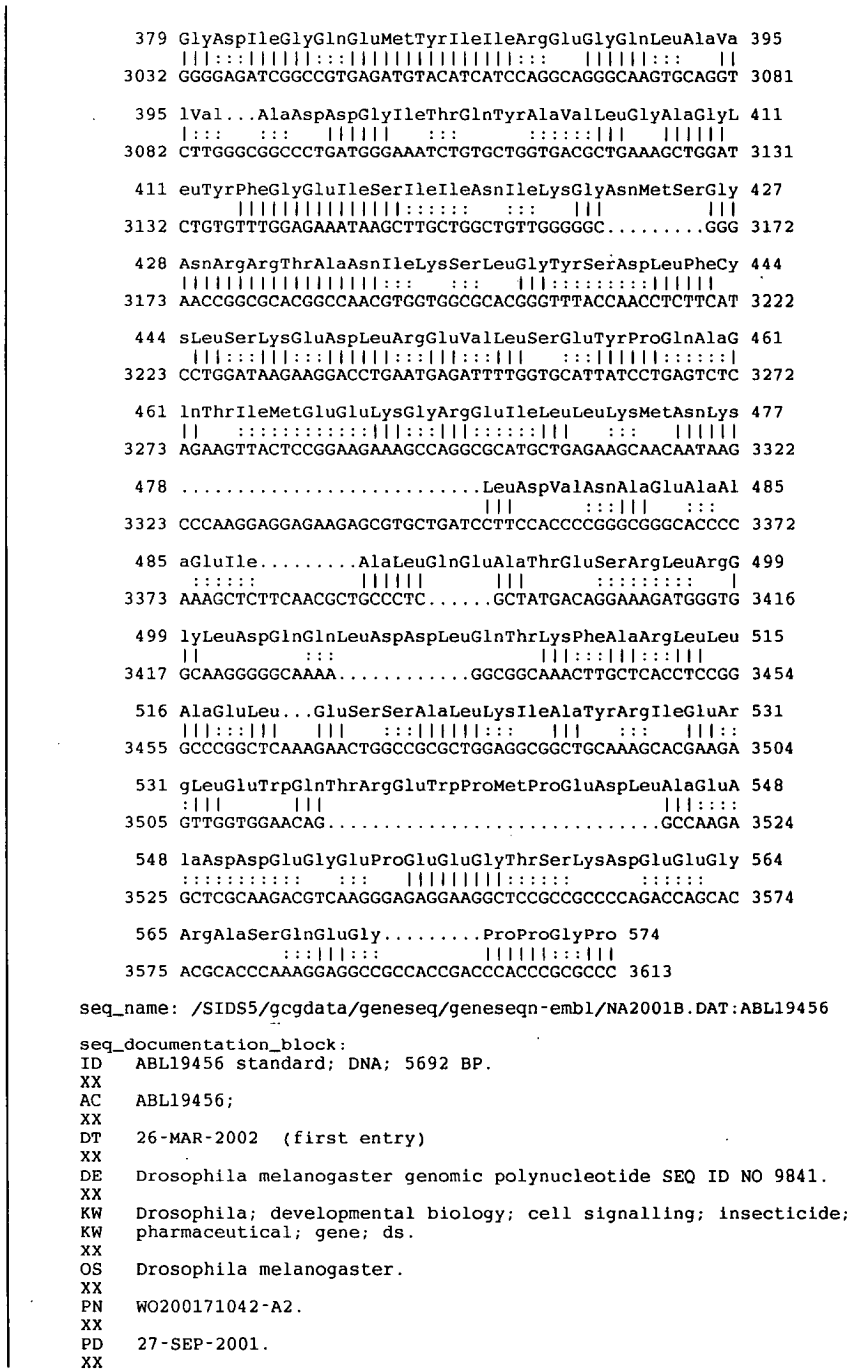
 Align seg 1/1  to: AAS91658  from: 1  to: 3811

      34 TyrTyrTrpTrpLeuAsnThrMetValPheProValMetTyrAsnLeuIl 50
         |||      ||||||      :::|||:::          :::|||
    2021 TATGTCCTATGGCTGTTCTTCGTGGTGATGGCCTGGAATTGGAACTGTTG 2070

      50 eIleLeuValCysArgAlaCysPheProAspLeuGlnHisGlyTyrLeuV 67
          ::::::      |||   :::|||||||                  :::
    2071 GCTGATTCCCGTGCGCTGGGCCTTCCCCTACCAGACCCCGGACAACATCC 2120

      67 alAlaTrpLeuValLeuAspTyrThrSerAspLeuLeuTyrLeuLeuAsp 83
             ||||||::::::||||||      ||||||:::|||:::||||||
```





```
seq_name: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL19456

seq_documentation_block:
ID   ABL19456 standard; DNA; 5692 BP.
XX
AC   ABL19456;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 9841.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
```

```
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 9841; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 5692 BP; 1433 A; 1294 C; 1301 G; 1664 T; 0 other;


alignment_scores:
            Quality:  695.00              Length:    678
              Ratio:  1.834                Gaps:     21
 Percent Similarity:  55.900   Percent Identity: 29.499

alignment_block:
 US-09-927-267-1 x ABL19456/rev   .:

 Align seg 1/1  to reverse of: ABL19456  from: 1  to: 5692

      10 ThrGluSerSerProProAlaProSerLysAlaArgLysLeuLeuPro.. 25
         |||:::|||      |||:::|||:::          :::|||:::|||
    2382 ACCAAGAGTCTGCACCCTTCGCCGCAGCACAAGGTGCGTCTGGTGCCGGC 2333

      25 ..................................................... 25

    2332 TGGTCAATCACCCCATCGTGGAAGACTCTTCGAAGCGGATACTCCAAGGT 2283

      25 ..................................................... 25

    2282 CCAATGTGTGGCTCTGTTCCAGTTTGTGCGGCGCCAACAATGATGAGCGA 2233

      26 ValLeuAspProSerGlyAspTyrTyrTyrTrpTrpLeuAsnThrMetVa 42
         :::||||||||||:::|||      |||      ||||||   ::::::
    2232 ACCCTGGATCCCCAGGGCAAGATCTACATCTCCTGGCTATGCGTTGTATC 2183

      42 lPheProValMetTyrAsnLeuIleIleLeuValCysArgAlaCysPheP 59
          :::      :::||||||      ::::::      ||||||   ||||
    2182 GCTATCGTTTCTCTATAACGCCTGGGTAATCCCGCTGCGTGCCTCGTTTC 2133

      59 roAspLeuGlnHisGlyTyrLeuValAlaTrpLeuValLeuAspTyrThr 75
         ||                          ||||||:::   |||:::
    2132 CCTTTCAAACCAAGGAGAACACCAATATCTGGCTGGCCTGCGACTTTTGC 2083

      76 SerAspLeuLeuTyrLeuLeuAspMetValValArgPheHis...ThrGl 91
         :::|||::::::||||||||||||:::|||      |||   :::
    2082 GCTGACATCATATATCTACTGGATGTGGTCTTCTTTAAACATCGGGTTAT 2033

      91 yPheLeuGluGlnGlyIleLeuValValAspLysGlyArgIleSerSerA 108
          :::|||   :::|||:::   |||   :::|||:::         ::::
    2032 GTACCTTTTTGAGGGATTCTGGGTGAAAAACAAGAACCTCACCAGAAAGA 1983

     108 rgTyrValArgThrTrpSerPhe..................... 115
         ::|||:::|||      :::|||
    1982 ACTACATGCGAAAACTGCAGTTTAAGGTGAGTTAACTTACAATTTCATTT 1933

     116 ........................Phe..LeuAspLeuAlaSerLeu 122
                                 |||  ||||||||   :::|||
    1932 AGAATTGTACAACTTAACTTAAAATGTTTCAGCTGGATCTGCTGGCCCTT 1883

     123 MetProThrAspValValTyrValArgLeuGlyProHisThrProThrLe 139
         :::|||   ::::::::::|||   :::||||||   ::::::      ||
    1882 CTTCCGCTGGAGCTATTGTACTTTAAACTCGGAACACAGGCGGTTTGGTT 1833

     139 uArgLeuAsnArgPheLeuArgAlaProArgLeuPheGluAlaPhe.... 154
         ||||:::   ||||||::::::         ::::::|||:::|||
    1832 GCGTTTTCCAAGGTTCTTTAAGATTCAGAGCTTCTGGGAAGTGTTTCGCC 1783

     155 .....AspArgThrGluThrArg.................ThrAlaTyr 163
               ||||||:::   :::                   |||   |||
    1782 TTTTGGATCGAGTCATATCTTCTCCGCACTTTGTAAGATATACTAATTAC 1733

     164 ProAsn..Ala..............................PheArg 168
            |||  :::                               |||
    1732 TTAAACTATGCCCTGTTTATTAACATACAATTTCAATATTATAGGTTCGA 1683

     169 IleAlaLysLeuMetLeuTyrIlePheValValIleHisTrpAsnSerCy 185
         :::||||||   :::   |||::::::   :::||||||   ::::::::
    1682 GTGGCCAAAACCCTAACCTACATGCTCTACATGATCCACATCACGGCTGC 1633

     185 sLeuTyrPheAlaLeuSerArgTyrLeuGlyPheGlyArgAspAlaTrpV 202
         :||||||:::|||   |||   |||   |||:::|||::::::   ||||
    1632 TCTGTACTACGCCTATAGTGACTACCAGGGTCTGGGACAGAATCGTTGGG 1583

     202 alTyrProAspProAlaGlnProGlyPheGluArgLeuArgArgGlnTyr 218
         ||:::      ::::::|||                          |||
    1582 TCTTCAGCGGCAAAGGGCATCCC.......................TAT 1557

     219 LeuTyrSerPheTyrPheSerThrLeuIleLeuThrThrValGlyAspTh 235
         :::      |||   |||:::|||         |||:::::::|||   ::
    1556 GTAAGGTGCTTTGCTTTTGCCACCAAGACGGCAACTTCGATAGGAAAAAA 1507

     235 rProProProAlaArgGluGluGluTyrLeuPheMetValGlyAspPheL 252
         :|||   |||   |||:::   ||||||:::||||||:::      :::|
    1506 CCCCAAGCCGGAACGTCAGGGAGAGTACGTCTTCATGACGGTGGCCTGGC 1457

     252 euLeuAlaValMetGlyPheAlaThrIleMetGlySerMetSerSerVal 268
         ||:::::::|||:::   ||||||   ::::::|||::::::   ::::::
    1456 TAATGGGCGTCTTTGTGTTCGCTCTGCTTATCGGTCAGATCAGGGATATC 1407

     269 Ile......TyrAsnMetAsnThrAlaAspAlaAlaPheTyrProAspHi 283
         |||         :::   |||      :::                 |||::
    1406 ATCTCGACAGCCACTCGGAATAAGCATGAGTATCGCCAGCTGGAGGATGA 1357

     283 sAlaLeuValLysLysTyrMetLysLeuGlnHisValAsnArgLysLeuG 300
         ::::|||      :::||||||:::     ::::::::::::::::::
    1356 GACACTG......GAGTACATGCGACGCCTCAATCTCTCGCAGGAGGTGC 1313

     300 luArgArgValIleAspTrpTyrGln..................... 308
         ::   ||||||     |||:::|||
    1312 AGTCGAGGGTCAAAATGTGGTTCCAGTTTACATGGGAGCAGCAGCGCACT 1263

     309 ............................HisLeuGln.IleAsnL 314
                                     :::   ||| |||:::
    1262 TTGGGTTTGTAAACTTCTAATATATGTTATATGCAATATCAAAATAAGAT 1213

     314 ysLysMetThrAsnGluValAlaIleLeuGlnHisLeuProGluArgLeu 330
           ::::::::::::|||      ||||||:::   ||||||   :::|||
    1212 TAAATTTTGCAGATGAATCCAACATATTGGATGCTTTGCCCATCAATTTG 1163

     331 ArgAlaGluValAlaValSerValHisLeuSerThrLeuSerArgValGl 347
         ::::::::::::|||:::|||||||||:::::|||||||||:::|||||
    1162 AAAACCGATATCGCCATCTCCGTGCACATCCAAACGCTCTCCAAGGTGCA 1113
```

```
     347 nIlePheGlnAsnCysGluAlaSerLeuLeuGluGluLeuValLeuLysL 364
         |:::|||   :::|||||||   :::|||||||:::::::||||||||||||
    1112 ACTCTTTGCTGATTGCGAGGAGGCTTTACTCAGAGATCTGGTTCTAAAAC 1063

     364 euGlnProGlnThrTyrSerProGlyGluTyrValCysArgLysGlyAsp 380
         ||:::      |||:::   |||||:::::::|||||||||||||||:::
    1062 TAAGGGCAGTCACATTTTTGCCAGGAGATTTCGTTTGCCGCAAAGGCGAA 1013

     381 IleGlyGlnGluMetTyrIleIleArgGluGlyGlnLeuAlaValVal.. 396
         :::|||:::|||||||||||||:::::::   ||||||:::   |||:::
    1012 GTGGGCAGAGAGATGTATATCGTGAAACTGGGACAGGTTCAAGTGATGGG 963

     397 .......AlaAspAspGlyIleThrGlnTyrAlaValLeuGlyAlaGlyL 411
               :::::::|||   :::         |||:::|||      |||
     962 TGGTCCTAGTAGCGATGTAGTG......CTGGCCACTCTAACCGAAGGTT 919

     411 euTyrPheGlyGluIleSerIleIleAsnIleLysGlyAsnMetSerGly 427
           |||||||||||||||:::::::::|||:::|||         :::
     918 CCGTGTTCGGGGAGATCAGTCTGCTGGGAATAAATGGA.........GCC 878

     428 AsnArgArgThrAlaAsnIleLysSerLeuGlyTyrSerAspLeuPheCy 444
         :::|||||||||||||:::::::::|||   |||||||||:::|||||||
     877 GATCGCAGAACAGCAGATGTCCGATCCAAGGGCTACTCCAACCTGTTCGT 828

     444 sLeuSerLysGluAspLeuArgGluValLeuSerGluTyrProGlnAlaG 461
          |||||||||:::|||||||:::|||||||::::::   ||||||   ||||
     827 GCTGTCCAAATCGGATTTAAACGAGGTCATTGCCTACTATCCCACGGCCC 778

     461 lnThrIleMetGluGluLysGlyArgGluIleLeuLeuLys......Met 475
         ||:::|||:::::::::::::::|||:::::::::   |||
     777 AGGCGATTCTCAAAAAGCGTGCTCGTCAGCTAATGCGAAAAAATGCAGCC 728

     476 AsnLysLeuAspValAsnAlaGluAlaAlaGluIleAlaLeuGln..... 490
         ::::::   :::   :::   |||   |||:::   |||||||||
     727 AGAGAGCGGGAAGAGGAGCGAGAGCGAGCACGCAGCGCCCTCCAGGCAGA 678

     491 .........................GluAlaThrGluSerArgLeuArgG 499
                                    |||      :::|||
     677 TGTGGTTATCGGCAATCCCAAAACACCAGAGACGGCGCCCAAACTT.... 632

     499 lyLeuAspGlnGlnLeuAspAspLeuGlnThrLysPheAlaArgLeuLeu 515
                              |||||||||   :::      |||
     631 .....................CTGCAGACAGTTATCCAGGCGCTG... 608

     516 AlaGluLeuGluSerSerAlaLeuLysIleAlaTyrArgIleGluArgLe 532
           :::||||||   |||:::  :::      |||   ::::::
     607 ...CCGTTTGAATCGCCCGCCGTGGTGCTCATCACGAGAGGTTCGAAGAG 561

     532 uGluTrpGlnThrArgGluTrpProMetProGluAspLeuAlaGluAlaA 549
              :::   |||:::      :::   |||   :::::::|||
     560 GATGCGCCGCAAGCGCCAGTCCGTGCAGATGGAGACCATTGTTGAGCCCA 511

     549 spAspGluGlyGluProGluGluGlyThrSerLysAspGluGluGlyArg 565
           |||               |||:::::::::   ::::::   :::
     510 AGATGGAG.........GTGACAGGCGCTGCCGAGACGCAAGATCAAAAG 470

     566 AlaSerGlnGluGlyProProGlyProGlu 575
         :::   :::   |||      :::|||:::
     469 GGCTTAAAGACAGGACGATGTTCGCCAGAT 440

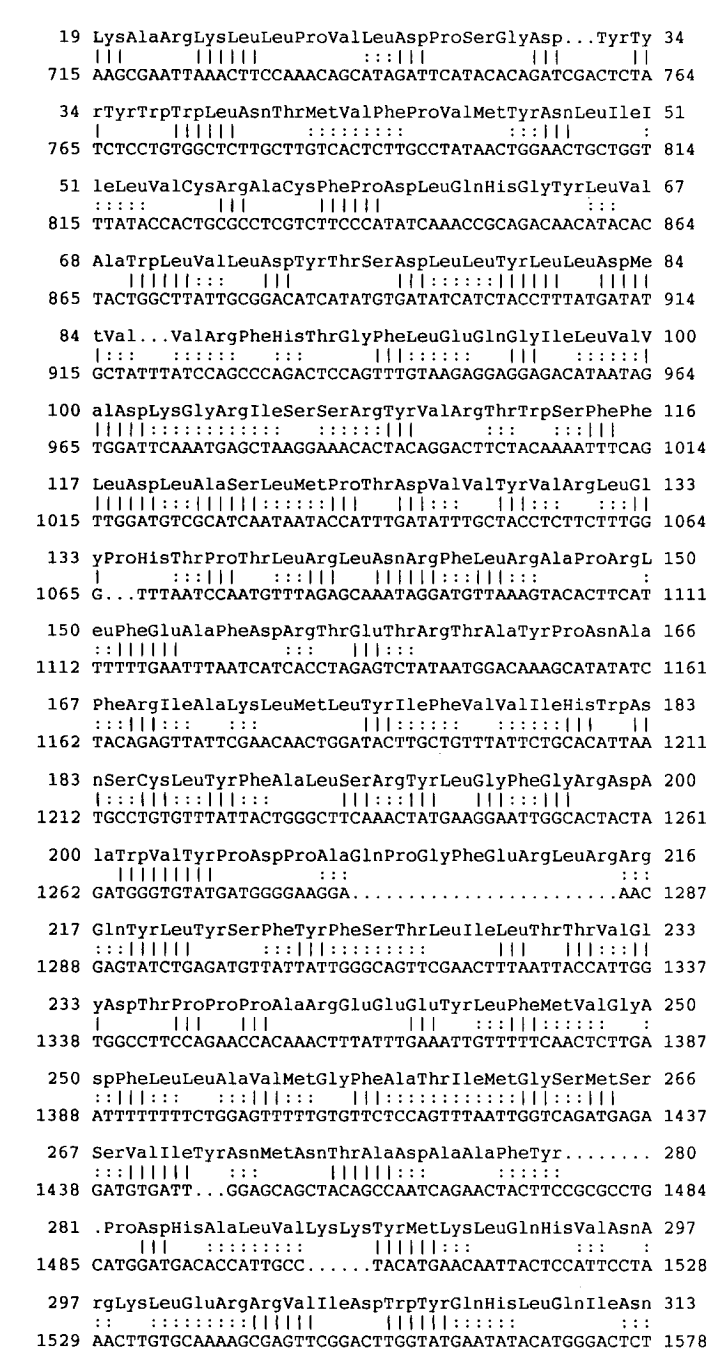
seq_name: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:AAD25729

seq_documentation_block:
ID   AAD25729 standard; DNA; 2757 BP.
XX
AC   AAD25729;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human cyclic nucleotide-gated cation channel 3 beta subunit DNA.
XX
KW   Human; cyclic nucleotide-gated cation channel 3 beta subunit; CNG3B;
KW   cyclic nucleotide-gated ion channel; contraceptive; vision disorder;
KW   male infertility; genetic defect; reporter-ligand interaction; CNG;
KW   viral infection; cancer; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1..49
FT                   /*tag= a
FT                   /note= "These bases are absent in the sequence shown
FT                   as SEQ ID NO:2 in Page 64-65 of the specification"
FT   CDS             112..2541
FT                   /*tag= b
FT                   /product= "Human CNG3B protein"
FT                   /note= "This region is specifically claimed in
FT                   claim 4 as SEQ ID NO:4"
FT   misc_feature    2753..2757
FT                   /*tag= c
FT                   /note= "These bases are absent in the sequence shown
FT                   as SEQ ID NO:2 in Page 64-65 of the specification"
XX
PN   WO200188090-A2.
XX
PD   22-NOV-2001.
XX
PF   15-MAY-2001; 2001WO-US15814.
XX
PR   15-MAY-2000; 2000US-204445P.
PR   14-MAY-2001; 2001US-0855828.
XX
PA   (ICAG-) ICAGEN INC.
XX
PI   Creech CD,  Jegla TJ;
XX
DR   WPI; 2002-089847/12.
DR   P-PSDB; AAE15982.
XX
PT   New polypeptide, useful for screening for modulators of cyclic
PT   nucleotide-gated ion channels, comprises the isolated cyclic
PT   nucleotide-gated cation channel 3 beta subunit -
XX
PS   Claim 4; Fig 2; 83pp; English.
XX
CC   The invention relates to human cyclic nucleotide-gated cation channel
CC   3 beta subunit (CNG3B) polypeptides and polynucleotides. CNG3B is a
CC   member of cyclic nucleotide-gated cation channel (CNG) family. CNG3B
CC   forms functional heteromultimers with CNGA1 and/or CNGA3, two CNG alpha
CC   subunits that are expressed in retina. CNG3B polypeptides are useful
CC   for screening modulators of CNGs which are useful as contraceptives and
CC   for treating various disorders involving cation channels, e.g. vision
CC   disorders and male infertility. Polynucleotides of the invention are
CC   useful for transfection of cells in vitro and in vivo, to correct
CC   acquired and inherited genetic defects, cancer and viral infections.
CC   Sequences of the invention are useful as reporter molecules in assays
CC   and detection systems, to measure changes in cation concentration,
CC   membrane potential, current flow, ion flux, transcription, signal
CC   transduction, reporter-ligand interactions and second messenger
CC   concentrations, in vitro, in vivo and ex vivo. They are useful to
CC   construct models of CNGs in a computer system and for examining
CC   expression and regulation of cation channels. The present sequence
CC   is a DNA encoding CNG3B protein.
XX
SQ   Sequence 2757 BP; 946 A; 571 C; 581 G; 659 T; 0 other;


alignment_scores:
           Quality:  655.00              Length:    565
             Ratio:  1.915                 Gaps:     14
 Percent Similarity:  60.531   Percent Identity: 29.204

alignment_block:
 US-09-927-267-1 x AAD25729   ..

 Align seg 1/1  to: AAD25729  from: 1  to: 2757
```



seq_name: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS91258

seq_documentation_block:
ID AAS91258 standard; cDNA; 1125 BP.
XX
AC AAS91258;
XX
DT 13-FEB-2002 (first entry)
XX

```
DE   DNA encoding novel human diagnostic protein #27062.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG27071.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 27062; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1125 BP; 283 A; 276 C; 228 G; 338 T; 0 other;


alignment_scores:
            Quality:  571.00          Length:    203
              Ratio:  3.461            Gaps:      0
 Percent Similarity:  81.281   Percent Identity: 50.739

alignment_block:
 US-09-927-267-1 x AAS91258   ..

 Align seg 1/1  to: AAS91258  from: 1  to: 1125

      92 PheLeuGluGlnGlyIleLeuValValAspLysGlyArgIleSerSerAr 108
         :::|||:::||||||:::||||||      :::   ::::::   :::::
     517 TACTTAAACCAAGGACTGTTGGTAAAGAAAGAACTTAAACTCATAAATAA 566

     108 gTyrValArgThrTrpSerPhePheLeuAspLeuAlaSerLeuMetProT 125
         :|||      :::   :::|||   |||||||:::   |||||||:::||||
     567 ATATAAATCCAACTTGCAATTTAAACTTGATGTTCTGTCACTGATACCAA 616

     125 hrAspValValTyrValArgLeuGlyProHisThrProThrLeuArgLeu 141
         |||||:::::|||   :::||||||   :::   |||   :::|||||||
     617 CTGATTTGCTGTATTTTAAGTTAGGGTGGAACTATCCAGAAATTAGATTA 666

     142 AsnArgPheLeuArgAlaProArgLeuPheGluAlaPheAspArgThrGl 158
         ||||||:::||||||      |||:::||||||   |||:::||||||||
     667 AACAGGTTGTTACGGTTCTCTCGTATGTTTGAGTTCTTCCAGAGAACAGA 716

     158 uThrArgThrAlaTyrProAsnAlaPheArgIleAlaLysLeuMetLeuT 175
         ||||||||||   ||||||||   |||||||||:::::|||:::::::|
     717 AACAAGGACAAACTATCCAAACATCTTCAGGATTTCCAACCTTGTTATGT 766

     175 yrIlePheValValIleHisTrpAsnSerCysLeuTyrPheAlaLeuSer 191
         |||||   ::::::||||||||||||:::|||:::::::::::::::|||
     767 ATATCGTCATCATTATCCACTGGAATGCATGTGTGTTCTACTCTATTTCT 816

     192 ArgTyrLeuGlyPheGlyArgAspAlaTrpValTyrProAspProAlaGl 208
         :::   :::|||||||||:::|||:::|||||||||||||||      ::
     817 AAAGCTATTGGATTTGGAAATGATACATGGGTCTACCCTGATATTAATGA 866

     208 nProGlyPheGluArgLeuArgArgGlnTyrLeuTyrSerPheTyrPheS 225
         :|||   |||   ||||||   |||:::|||:::|||||:::|||:::|
     867 TCCTGAATTTGGCCGTTTGGCTAGAAAATACGTATACAGCCTTTACTGGT 916

     225 erThrLeuIleLeuThrThrValGlyAspThrProProProAlaArgGlu 241
         ||||||||   |||||||||:::|||:::|||||||||||||:::|||:::
     917 CTACACTGACTTTGACTACCATTGGTGAAACACCCCCTCCCGTGAGGGAT 966

     242 GluGluTyrLeuPheMetValGlyAspPheLeuLeuAlaValMetGlyPh 258
         :::||||||:::|||:::|||   ||||||||:::::::|||::   ||
     967 TCTGAGTATGTCTTTGTGGTGGTTGATTTCCTAATTGGAGTGTTAATTTT 1016

     258 eAlaThrIleMetGlySerMetSerSerValIleTyrAsnMetAsnThrA 275
         ||||||||||:::|||:::::::::|||:::|||   |||||||||:::|
    1017 TGCTACCATCGTTGGTAACATAGGTTCTATGATTTCCAACATGAATGCAG 1066

     275 laAspAlaAlaPheTyrProAspHisAlaLeuValLysLysTyrMetLys 291
         ||   |||   |||                  :::|||:::||||||
    1067 CCAGAGCAGAATTTCAAGCAAGAATTGATGCTATCAAGCAATATATGCAT 1116

     292 LeuGlnHis 294
         :::::::::
    1117 TTTCGAAAT 1125

seq_name: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS79945

seq_documentation_block:
ID   AAS79945 standard; cDNA; 1146 BP.
XX
AC   AAS79945;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #15749.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
```

```
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG15758.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 15749; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1146 BP; 357 A; 219 C; 266 G; 304 T; 0 other;


alignment_scores:
            Quality:  540.00              Length:    216
              Ratio:  3.396                Gaps:      2
 Percent Similarity:  73.611  Percent Identity: 48.148

alignment_block:
 US-09-927-267-1 x AAS79945   ..

 Align seg 1/1  to: AAS79945  from: 1  to: 1146

     241 GluGluGluTyrLeuPheMetValGlyAspPheLeuLeuAlaValMetGl 257
         ::::::|||||:::|||:::|||   ||||||||:::::|||:::
       1 GATTCTGAGTATGTCTTTGTGGTGGTTGATTTCCTAATTGGAGTGTTAAT 50

     257 yPheAlaThrIleMetGlySerMetSerSerValIleTyrAsnMetAsnT 274
          ||||||||||||:::|||:::::::::|||:::|||   |||||||||:
      51 TTTTGCTACCATCGTTGGTAACATAGGTTCTATGATTTCCAACATGAATG 100

     274 hrAlaAspAlaAlaPheTyrProAspHisAlaLeuValLysLysTyrMet 290
         ::|||   |||   |||                 :::|||:::|||:::
     101 CAGCCAGAGCAGAATTTCAAGCAAGAATTGATGCTATCAAGCAATATGTG 150

     291 Lys.....................................LeuGl 293
         :::                                     :::::
     151 CGTGGTCGAAATCAAGATTTGATGCTATCAAGCAATATATGCACTTTTCG 200

     293 nHisValAsnArgLysLeuGluArgArgValIleAspTrpTyrGlnHisL 310
         ::::|||::::::   :::|||:::|||||||||   |||:::::::::|
     201 AAATGTAAGCAAAGATATGGAAAAGAGGGTTATTAAATGGTTTGACTACC 250

     310 euGlnIleAsnLysLysMetThrAsnGluValAlaIleLeuGlnHisLeu 326
         ||      |||||||||   ::::::|||      :::|||:::::::|||
     251 TGTGGACCAACAAAAAAACAGTTGATGAGAAAGAAGTCTTAAAGTATCTA 300

     327 ProGluArgLeuArgAlaGluValAlaValSerValHisLeuSerThrLe 343
         |||::::::|||||||||||:::|||::::::|||||||||:::|||::
     301 CCTGATAAACTAAGAGCAGAAATTGCCATCAACGTTCACTTAGACACAGT 350

     343 uSerArgValGlnIlePheGlnAsnCysGluAlaSerLeuLeuGluGluL 360
         :::::::|||
     351 GGGGAAAGTC..................................... 360

     360 euValLeuLysLeuGlnProGlnThrTyrSerProGlyGluTyrValCys 376
                                     ||||||||||||:::|||:::|||
     361 .........................TACAGTCCTGGAGATTATATTTGC 384

     377 ArgLysGlyAspIleGlyGlnGluMetTyrIleIleArgGluGlyGlnLe 393
         :::|||||||||||||||:::|||||||||||||||:::||||||:::||
     385 AAGAAAGGGGATATCGGACGAGAGATGTACATTATCAAGGAAGGCAAACT 434

     393 uAlaValValAlaAspAspGlyIleThrGlnTyrAlaValLeuGlyAlaG 410
         ||||||||||||||||||||||:::||||||:::::||||||:::   |
     435 CGCTGTGGTGGCAGATGATGGAGTCACTCAGTTTGTGGTATTGAGCGATG 484

     410 lyLeuTyrPheGlyGluIleSerIleIleAsnIleLysGlyAsnMetSer 426
         ||   |||||||||||||||||||||:::|||||||||||||:::   :::
     485 GCAGCTACTTCGGTGAGATCAGCATTCTTAACATTAAAGGGAGCAAAGCT 534

     427 GlyAsnArgArgThrAlaAsnIleLysSerLeuGlyTyrSerAspLeu 442
         ||||||||||||||||||||||||||||:::|||||||||||||||||
     535 GGCAATCGAAGAACGGCCAATATTAAAAGTATTGGCTACTCAGACCTG 582

seq_name: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS77668

seq_documentation_block:
ID   AAS77668 standard; cDNA; 2421 BP.
XX
AC   AAS77668;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #13472.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG13481.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 13472; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
```

CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2421 BP; 715 A; 553 C; 591 G; 562 T; 0 other;

alignment_scores:
Quality: 540.00 Length: 216
Ratio: 3.396 Gaps: 2
Percent Similarity: 73.611 Percent Identity: 48.148

alignment_block:
US-09-927-267-1 x AAS77668 ..

Align seg 1/1 to: AAS77668 from: 1 to: 2421

```
241 GluGluGluTyrLeuPheMetValGlyAspPheLeuLeuAlaValMetGl 257
    ::::::||||||:::|||:::|||   |||||||||:::::::|||:::
  1 GATTCTGAGTATGTCTTTGTGGTGGTTGATTTCCTAATTGGAGTGTTAAT 50

257 yPheAlaThrIleMetGlySerMetSerSerValIleTyrAsnMetAsnT 274
     ||||||||||||:::|||:::::::::|||:::|||   |||||||||:
 51 TTTTGCTACCATCGTTGGTAACATAGGTTCTATGATTTCCAACATGAATG 100

274 hrAlaAspAlaAlaPheTyrProAspHisAlaLeuValLysLysTyrMet 290
    ::|||   |||   |||                :::|||:::|||:::
101 CAGCCAGAGCAGAATTTCAAGCAAGAATTGATGCTATCAAGCAATATGTG 150

291 Lys.....................................LeuGl 293
    :::                                     :::::
151 CGTGGTCGAAATCAAGATTTGATGCTATCAAGCAATATATGCACTTTTCG 200

293 nHisValAsnArgLysLeuGluArgArgValIleAspTrpTyrGlnHisL 310
    ::::|||:::::::   :::|||:::|||||||||   |||:::::::::|
201 AAATGTAAGCAAAGATATGGAAAAGAGGGTTATTAAATGGTTTGACTACC 250

310 euGlnIleAsnLysLysMetThrAsnGluValAlaIleLeuGlnHisLeu 326
    ||      |||||||||   ::::::|||      :::|||::::::|||
251 TGTGGACCAACAAAAAAACAGTTGATGAGAAAGAAGTCTTAAAGTATCTA 300

327 ProGluArgLeuArgAlaGluValAlaValSerValHisLeuSerThrLe 343
    |||:::::|||||||||||||:::|||:::::::|||||||||:::|||::
301 CCTGATAAACTAAGAGCAGAAATTGCCATCAACGTTCACTTAGACACAGT 350

343 uSerArgValGlnIlePheGlnAsnCysGluAlaSerLeuLeuGluGluL 360
    ::::::::|||
351 GGGGAAAGTC..................................... 360

360 euValLeuLysLeuGlnProGlnThrTyrSerProGlyGluTyrValCys 376
                              ||||||||||||:::|||:::|||
361 .........................TACAGTCCTGGAGATTATATTTGC 384

377 ArgLysGlyAspIleGlyGlnGluMetTyrIleIleArgGluGlyGlnLe 393
    :::|||||||||||||||:::|||||||||||||||:::||||||:::||
385 AAGAAAGGGGATATCGGACGAGAGATGTACATTATCAAGGAAGGCAAACT 434

393 uAlaValValAlaAspAspGlyIleThrGlnTyrAlaValLeuGlyAlaG 410
    ||||||||||||||||||||:::||||||:::::::||||||:::   |
435 CGCTGTGGTGGCAGATGATGGAGTCACTCAGTTTGTGGTATTGAGCGATG 484

410 lyLeuTyrPheGlyGluIleSerIleIleAsnIleLysGlyAsnMetSer 426
    ||   ||||||||||||||||||:::|||||||||||||:::   :::
485 GCAGCTACTTCGGTGAGATCAGCATTCTTAACATTAAAGGGAGCAAAGCT 534

427 GlyAsnArgArgThrAlaAsnIleLysSerLeuGlyTyrSerAspLeu 442
    ||||||||||||||||||||||||||||||:::|||||||||||||||
535 GGCAATCGAAGAACGGCCAATATTAAAAGTATTGGCTACTCAGACCTG 582
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2002, 09:08:57 ; Search time 7623.77 Seconds
(without alignments)
6335.244 Million cell updates/sec

Title: US-09-927-267-2
Perfect score: 2308
Sequence: 1 agaggggaggaggaaaacag.....agactgtgcattccatctaa 2308

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2101.8	91.1	2551	6	AX350404	AX350404 Sequence
2	2009.8	87.1	2013	6	AX355967	AX355967 Sequence
3	1785.4	77.4	1835	6	AX350402	AX350402 Sequence
4	1508.4	65.4	2351	10	RNU12623	U12623 Rattus norv
5	1470.2	63.7	1938	10	RNU12425	U12425 Rattus norv
6	687.2	29.8	164434	2	AC036216	AC036216 Homo sapi
7	684.4	29.7	162965	2	AC021935	AC021935 Homo sapi
8	684.4	29.7	186609	2	AC022762	AC022762 Homo sapi
9	617.4	26.8	2641	4	BTCNGCHAN	X89600 B.taurus mR
10	617.4	26.8	3998	4	BTCNG3	X76485 B.taurus CN
11	588.2	25.5	3486	9	AF065314	AF065314 Homo sapi
12	584.8	25.3	2781	10	MMU243933	AJ243933 Mus muscu
13	583.2	25.3	3166	4	BTCAMPGC	X55010 B.taurus mR
14	578.6	25.1	1995	6	AX179746	AX179746 Sequence
15	572	24.8	2669	4	OCRACNG	X59668 O.cuniculus
16	570	24.7	2098	10	RNO272429	AJ272429 Rattus no
17	570	24.7	2212	10	RNO272428	AJ272428 Rattus no
18	570	24.7	2213	10	AB002801	AB002801 Rattus no
19	530	23.0	2815	10	AF126808	AF126808 Rattus no
20	530	23.0	3027	10	RNOLFCH	X55519 Rat mRNA fo
21	515.2	22.3	153551	9	AC092675	AC092675 Homo sapi
22	511	22.1	1428	10	MMU238241	AJ238241 Mus muscu
23	507.4	22.0	2697	5	GGCONECNG	X89598 G.gallus mR
24	499.2	21.6	3718	5	AF393839	AF393839 Oncorhync
25	498.6	21.6	104037	9	HSAF002992	AF002992 Homo sapi
26	494.8	21.4	2049	5	ICTCNC	M83111 Ictalurus p
27	490	21.2	2052	10	RNU48803	U48803 Rattus norv
28	481.2	20.8	2148	10	RNU93851	U93851 Rattus norv
29	475.6	20.6	2682	4	BTCGMPCH	X51604 B.taurus RN
30	475.4	20.6	158588	2	AC091450	AC091450 Mus muscu
31	472.2	20.5	9618	10	MMU49391	U49391 Mus musculu
32	467.6	20.3	2697	4	CFACGMP	X99914 C.familiari
33	466	20.2	2297	10	MMU19717	U19717 Mus musculu
34	466	20.2	2717	4	CFU83905	U83905 Canis famil
35	464.4	20.1	2474	10	MUSCNCG	M84742 Mus musculu
36	461.2	20.0	2061	4	SSU85404	U85404 Sus scrofa
37	461.2	20.0	2857	9	S42457	S42457 CNCG=rod ph
38	459	19.9	2612	5	GGRODCNG	X89599 G.gallus mR
39	452.4	19.6	2500	9	HUMCGMP	M84741 Human cGMP-
40	423.2	18.3	200039	2	AL671880	AL671880 Mus muscu
c 41	423.2	18.3	218872	2	AC099698	AC099698 Mus muscu
42	416.8	18.1	222364	30	AC013535	AC013535 Homo sapi
c 43	416.8	18.1	229687	2	AC107068	AC107068 Homo sapi
44	410	17.8	2913	3	DMCNGCHAN	X89601 D.melanogas
c 45	386.4	16.7	190000	2	AC007704	AC007704 Homo sapi

ALIGNMENTS

RESULT 1
AX350404
LOCUS AX350404 2551 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 29 from Patent WO0181578.
ACCESSION AX350404
VERSION AX350404.1 GI:18616046
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Vernet,C.A., Fernandes,E.R., Gerlach,V., Shimkets,R.A.,
Malyankar,U.M., Boldog,F.L., Zerhusen,B.D., Spytek,K.A.,
Majumder,K., Tchernev,V.T., Padigaru,M., Patturajan,M.,
Burgess,C.E., Gangolli,E.A., Smithson,G., Rastelli,L.,
Macdougall,J.R., Taupier,R.J., Grosse,W.M. and Alsobrook,J.P.
TITLE Novel proteins and nucleic acids encoding same
JOURNAL Patent: WO 0181578-A 29 01-NOV-2001;
Curagen Corporation (US)
FEATURES Location/Qualifiers
source 1. .2551
/organism="Homo sapiens"

```
                     /db_xref="taxon:9606"
BASE COUNT      646 a    668 c    707 g    530 t
ORIGIN


  Query Match             91.1%;  Score 2101.8;  DB 6;  Length 2551;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 2103;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 agaggggaggaggaaaacagagacaagactcaggcttccctctgaggcatgcacccccac 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     447 AGAGGGGAGGAGGAAAACAGAGACAAGACTCAGGCTTCCCTCTGAGGCATGCACCCCCAC 506

Qy      61 cttctccagggatctcattagaggtgtttagctgggcaggtgtaagcccaggccctggga 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     507 CTTCTCCAGGGATCTCATTAGAGGTGTTTAGCTGGGCAGGTGTAAGCCCAGGCCCTGGGA 566

Qy     121 gacagggcagagtgctagagctagactgtctccacccccttcagtagcgctagctctggtt 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     567 GACAGGGCAGAGTGCTAGAGCTAGACTGTCTCCACCCCTTCAGTAGCGCTAGCTCTGGTT 626

Qy     181 gtgttgctaagagccccaaagacaaagaagtcacagcagaagcccaacagcagcctcctt 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     627 GTGTTGCTAAGAGCCCCAAAGACAAAGAAGTCACAGCAGAAGCCCAACAGCAGCCTCCTT 686

Qy     241 cagacagtcaggcactagtgcccaactccagaagtcccctacaggcagagagggtgtgga 300
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     687 CAGGCAGTCAGGCACTAGTGCCCAACTCCAGAAGTCCCCTACAGGCAGAGAGGGTGTGGA 746

Qy     301 catctcacaccccagcaccagaccacagaaccatgagccaggacaccaaagtgaagacaa 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     747 CATCTCACACCCCAGCACCAGACCACAGAACCATGAGCCAGGACACCAAAGTGAAGACAA 806

Qy     361 cagagtccagtcccccagccccatccaaggccaggaagttgctgcctgtcctggacccat 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     807 CAGAGTCCAGTCCCCCAGCCCCATCCAAGGCCAGGAAGTTGCTGCCTGTCCTGGACCCAT 866

Qy     421 ctggggattactactactggtggctgaacacaatggtcttcccagtcatgtataacctca 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     867 CTGGGGATTACTACTACTGGTGGCTGAACACAATGGTCTTCCCAGTCATGTATAACCTCA 926

Qy     481 tcatcctcgtgtgcagagcctgcttccccgacttgcagcacggttatctggtggcctggt 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     927 TCATCCTCGTGTGCAGAGCCTGCTTCCCCGACTTGCAGCACGGTTATCTGGTGGCCTGGT 986

Qy     541 tggtgctggactacacgagtgacctgctatacctactagacatggtggtgcgcttccaca 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     987 TGGTGCTGGACTACACGAGTGACCTGCTATACCTACTAGACATGGTGGTGCGCTTCCACA 1046

Qy     601 caggattcttggaacagggcatcctggtggtggacaagggtaggatctcgagtcgctacg 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1047 CAGGATTCTTGGAACAGGGCATCCTGGTGGTGGACAAGGGTAGGATCTCGAGTCGCTACG 1106

Qy     661 ttcgcacctggagtttcttcttggacctggcttccctgatgcccacagatgtggtctacg 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1107 TTCGCACCTGGAGTTTCTTCTTGGACCTGGCTTCCCTGATGCCCACAGATGTGGTCTACG 1166

Qy     721 tgcggctgggcccgcacacacccaccctgaggctgaaccgctttctccgcgcgccccgcc 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1167 TGCGGCTGGGCCCGCACACACCCACCCTGAGGCTGAACCGCTTTCTCCGCGCGCCCCGCC 1226

Qy     781 tcttcgaggccttcgaccgcacagagacccgcacagcttacccaaatgcctttcgcattg 840
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    1227 TCTTCGAGGCCTTCGACCGCGCAGAGACCCGCACAGCTTACCCAAATGCCTTTCGCATTG 1286

Qy     841 ccaagctgatgctttacatttttgtcgtcatccattggaacagctgcctatactttgccc 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1287 CCAAGCTGATGCTTTACATTTTTGTCGTCATCCATTGGAACAGCTGCCTATACTTTGCCC 1346

Qy     901 tatcccggtacctgggcttcgggcgtgacgcatgggtgtacccggaccccgcgcagcctg 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1347 TATCCCGGTACCTGGGCTTCGGGCGTGACGCATGGGTGTACCCGGACCCCGCGCAGCCTG 1406

Qy     961 gctttgagcgcctgcggcgccagtacctctatagcttttacttctccacgctgatactga 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1407 GCTTTGAGCGCCTGCGGCGCCAGTACCTCTATAGCTTTTACTTCTCCACGCTGATACTGA 1466

Qy    1021 ctacagtgggcgatacaccgccgccagccagggaagaagagtacctcttcatggtgggcg 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1467 CTACAGTGGGCGATACACCGCCGCCAGCCAGGGAAGAAGAGTACCTCTTCATGGTGGGCG 1526

Qy    1081 acttcctgctggccgtcatgggtttcgccaccatcatgggtagcatgagctctgtcatct 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1527 ACTTCCTGCTGGCCGTCATGGGTTTCGCCACCATCATGGGTAGCATGAGCTCTGTCATCT 1586

Qy    1141 acaacatgaacactgcagatgcggctttctacccagatcatgcactggtgaagaagtaca 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1587 ACAACATGAACACTGCAGATGCGGCTTTCTACCCAGATCATGCACTGGTGAAGAAGTACA 1646

Qy    1201 tgaagctgcagcacgtcaaccgcaagctggagcggcgagttattgactggtatcagcacc 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1647 TGAAGCTGCAGCACGTCAACCGCAAGCTGGAGCGGCGAGTTATTGACTGGTATCAGCACC 1706

Qy    1261 tgcagatcaacaagaagatgaccaacgaggtagccatcttacagcacttgcctgagcggc 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1707 TGCAGATCAACAAGAAGATGACCAACGAGGTAGCCATCTTACAGCACTTGCCTGAGCGGC 1766

Qy    1321 tgcgggcagaagtggctgtgtctgtgcacctgtccactctgagccgggtgcagatctttc 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1767 TGCGGGCAGAAGTGGCTGTGTCTGTGCACCTGTCCACTCTGAGCCGGGTGCAGATCTTTC 1826

Qy    1381 agaactgtgaggccagcctgctggaggagctggtgctgaagctgcagccccagacctact 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1827 AGAACTGTGAGGCCAGCCTGCTGGAGGAGCTGGTGCTGAAGCTGCAGCCCCAGACCTACT 1886

Qy    1441 caccaggtgaatatgtatgccgcaaaggagacattggccaagagatgtacatcatccgag 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1887 CACCAGGTGAATATGTATGCCGCAAAGGAGACATTGGCCAAGAGATGTACATCATCCGAG 1946

Qy    1501 agggtcaactggccgtggtggcagatgatggtatcacacagtatgctgtgctcggtgcag 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1947 AGGGTCAACTGGCCGTGGTGGCAGATGATGGTATCACACAGTATGCTGTGCTCGGTGCAG 2006

Qy    1561 ggctctactttggggagatcagcatcatcaacatcaaagggaacatgtctgggaaccgcc 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2007 GGCTCTACTTTGGGGAGATCAGCATCATCAACATCAAAGGGAACATGTCTGGGAACCGCC 2066

Qy    1621 gcacagccaacatcaagagcctaggttattcagacctattctgcctgagcaaggaggacc 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2067 GCACAGCCAACATCAAGAGCCTAGGTTATTCAGACCTATTCTGCCTGAGCAAGGAGGACC 2126

Qy    1681 tgcgggaggtgctgagcgagtatccacaagcacagaccatcatggaggagaaaggacgtg 1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2127 TGCGGGAGGTGCTGAGCGAGTATCCACAAGCACAGACCATCATGGAGGAGAAAGGACGTG 2186

Qy    1741 agatcctgctgaaaatgaacaagttggacgtgaatgctgaggcagctgagatcgccctgc 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2187 AGATCCTGCTGAAAATGAACAAGTTGGACGTGAATGCTGAGGCAGCTGAGATCGCCCTGC 2246

Qy    1801 aggaggccacagagtcccggctacgaggcctagaccagcagctggatgatctacagacca 1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2247 AGGAGGCCACAGAGTCCCGGCTACGAGGCCTAGACCAGCAGCTGGATGATCTACAGACCA 2306

Qy    1861 agtttgctcgcctcctggctgagctggagtccagcgcacttaagattgcttaccgcattg 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2307 AGTTTGCTCGCCTCCTGGCTGAGCTGGAGTCCAGCGCACTTAAGATTGCTTACCGCATTG 2366

Qy    1921 aacggctggagtggcagactcgagagtggccaatgcccgaggacctggctgaggctgatg 1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2367 AACGGCTGGAGTGGCAGACTCGAGAGTGGCCAATGCCCGAGGACCTGGCTGAGGCTGATG 2426

Qy    1981 acgagggtgagcctgaggagggaacttccaaagatgaagagggcagggccagccaggagg 2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2427 ACGAGGGTGAGCCTGAGGAGGGAACTTCCAAAGATGAAGAGGGCAGGGCCAGCCAGGAGG 2486

Qy    2041 gacccccaggtccagagtgaccccatccccatccccaggattcccacctcctagtgaatc 2100
```

```
Db   2487 GACCCCCAGGTCCAGAGTGACCCCATCCCCATCCCCAGGATTCCCACCTCCTAGTGAATC 2546

Qy   2101 cagag 2105
          |||||
Db   2547 CAGAG 2551


RESULT   2
AX355967
LOCUS       AX355967               2013 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 2 from Patent WO0202633.
ACCESSION   AX355967
VERSION     AX355967.1  GI:18620561
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Raumann,B.E., Sanjanwala,M.S., Tribouley,C.M. and Walia,N.K.
  TITLE     Transporters and ion channels
  JOURNAL   Patent: WO 0202633-A 2 10-JAN-2002;
            Incyte Genomics, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .2013
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /note="Incyte ID No: 6703242CB1"
BASE COUNT      472 a    576 c    557 g    408 t
ORIGIN


  Query Match           87.1%;  Score 2009.8;  DB 6;  Length 2013;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 2011;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy    210 gtcacagcagaagcccaacagcagcctccttcagacagtcaggcactagtgcccaactcc 269
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db      1 GTCACAGCAGAAGCCCAACAGCAGCCTCCTTCAGGCAGTCAGGCACTAGTGCCCAACTCC 60

Qy    270 agaagtcccctacaggcagagagggtgtggacatctcacaccccagcaccagaccacaga 329
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AGAAGTCCCCTACAGGCAGAGAGGGTGTGGACATCTCACACCCCAGCACCAGACCACAGA 120

Qy    330 accatgagccaggacaccaaagtgaagacaacagagtccagtcccccagccccatccaag 389
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ACCATGAGCCAGGACACCAAAGTGAAGACAACAGAGTCCAGTCCCCCAGCCCCATCCAAG 180

Qy    390 gccaggaagttgctgcctgtcctggacccatctggggattactactactggtggctgaac 449
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GCCAGGAAGTTGCTGCCTGTCCTGGACCCATCTGGGGATTACTACTACTGGTGGCTGAAC 240

Qy    450 acaatggtcttcccagtcatgtataacctcatcatcctcgtgtgcagagcctgcttcccc 509
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ACAATGGTCTTCCCAGTCATGTATAACCTCATCATCCTCGTGTGCAGAGCCTGCTTCCCC 300

Qy    510 gacttgcagcacggttatctggtggcctggttggtgctggactacacgagtgacctgcta 569
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GACTTGCAGCACGGTTATCTGGTGGCCTGGTTGGTGCTGGACTACACGAGTGACCTGCTA 360

Qy    570 tacctactagacatggtggtgcgcttccacacaggattcttggaacagggcatcctggtg 629
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TACCTACTAGACATGGTGGTGCGCTTCCACACAGGATTCTTGGAACAGGGCATCCTGGTG 420

Qy    630 gtggacaagggtaggatctcgagtcgctacgttcgcacctggagtttcttcttggacctg 689
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GTGGACAAGGGTAGGATCTCGAGTCGCTACGTTCGCACCTGGAGTTTCTTCTTGGACCTG 480

Qy    690 gcttccctgatgcccacagatgtggtctacgtgcggctgggcccgcacacacccaccctg 749
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GCTTCCCTGATGCCCACAGATGTGGTCTACGTGCGGCTGGGCCCGCACACACCCACCCTG 540

Qy    750 aggctgaaccgctttctccgcgcgccccgcctcttcgaggccttcgaccgcacagagacc 809
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 AGGCTGAACCGCTTTCTCCGCGCGCCCCGCCTCTTCGAGGCCTTCGACCGCACAGAGACC 600

Qy    810 cgcacagcttacccaaatgcctttcgcattgccaagctgatgctttacatttttgtcgtc 869
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CGCACAGCTTACCCAAATGCCTTTCGCATTGCCAAGCTGATGCTTTACATTTTTGTCGTC 660

Qy    870 atccattggaacagctgcctatactttgccctatcccggtacctgggcttcgggcgtgac 929
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 ATCCATTGGAACAGCTGCCTATACTTTGCCCTATCCCGGTACCTGGGCTTCGGGCGTGAC 720

Qy    930 gcatgggtgtacccggaccccgcgcagcctggctttgagcgcctgcggcgccagtacctc 989
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GCATGGGTGTACCCGGACCCCGCGCAGCCTGGCTTTGAGCGCCTGCGGCGCCAGTACCTC 780

Qy    990 tatagcttttacttctccacgctgatactgactacagtgggcgatacaccgccgccagcc 1049
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    781 TATAGCTTTTACTTCCCCACGCTGATACTGACTACAGTGGGCGATACACCGCCGCCAGCC 840

Qy   1050 agggaagaagagtacctcttcatggtgggcgacttcctgctggccgtcatgggtttcgcc 1109
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 AGGGAAGAAGAGTACCTCTTCATGGTGGGCGACTTCCTGCTGGCCGTCATGGGTTTCGCC 900

Qy   1110 accatcatgggtagcatgagctctgtcatctacaacatgaacactgcagatgcggctttc 1169
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 ACCATCATGGGTAGCATGAGCTCTGTCATCTACAACATGAACACTGCAGATGCGGCTTTC 960

Qy   1170 tacccagatcatgcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctg 1229
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 TACCCAGATCATGCACTGGTGAAGAAGTACATGAAGCTGCAGCACGTCAACCGCAAGCTG 1020

Qy   1230 gagcggcgagttattgactggtatcagcacctgcagatcaacaagaagatgaccaacgag 1289
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 GAGCGGCGAGTTATTGACTGGTATCAGCACCTGCAGATCAACAAGAAGATGACCAACGAG 1080

Qy   1290 gtagccatcttacagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtgcac 1349
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 GTAGCCATCTTACAGCACTTGCCTGAGCGGCTGCGGGCAGAAGTGGCTGTGTCTGTGCAC 1140

Qy   1350 ctgtccactctgagccgggtgcagatctttcagaactgtgaggccagcctgctggaggag 1409
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CTGTCCACTCTGAGCCGGGTGCAGATCTTTCAGAACTGTGAGGCCAGCCTGCTGGAGGAG 1200

Qy   1410 ctggtgctgaagctgcagccccagacctactcaccaggtgaatatgtatgccgcaaagga 1469
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 CTGGTGCTGAAGCTGCAGCCCCAGACCTACTCACCAGGTGAATATGTATGCCGCAAAGGA 1260

Qy   1470 gacattggccaagagatgtacatcatccgagagggtcaactggccgtggtggcagatgat 1529
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 GACATTGGCCAAGAGATGTACATCATCCGAGAGGGTCAACTGGCCGTGGTGGCAGATGAT 1320

Qy   1530 ggtatcacacagtatgctgtgctcggtgcagggctctactttggggagatcagcatcatc 1589
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 GGTATCACACAGTATGCTGTGCTCGGTGCAGGGCTCTACTTTGGGGAGATCAGCATCATC 1380

Qy   1590 aacatcaaagggaacatgtctgggaaccgccgcacagccaacatcaagagcctaggttat 1649
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 AACATCAAAGGGAACATGTCTGGGAACCGCCGCACAGCCAACATCAAGAGCCTAGGTTAT 1440

Qy   1650 tcagacctattctgcctgagcaaggaggacctgcgggaggtgctgagcgagtatccacaa 1709
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 TCAGACCTATTCTGCCTGAGCAAGGAGGACCTGCGGGAGGTGCTGAGCGAGTATCCACAA 1500

Qy   1710 gcacagaccatcatggaggagaaaggacgtgagatcctgctgaaaatgaacaagttggac 1769
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 GCACAGACCATCATGGAGGAGAAAGGACGTGAGATCCTGCTGAAAATGAACAAGTTGGAC 1560

Qy   1770 gtgaatgctgaggcagctgagatcgccctgcaggaggccacagagtcccggctacgaggc 1829
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 GTGAATGCTGAGGCAGCTGAGATCGCCCTGCAGGAGGCCACAGAGTCCCGGCTACGAGGC 1620

Qy   1830 ctagaccagcagctggatgatctacagaccaagtttgctcgcctcctggctgagctggag 1889
```

```
Db    1621 CTAGACCAGCAGCTGGATGATCTACAGACCAAGTTTGCTCGCCTCCTGGCTGAGCTGGAG 1680

Qy    1890 tccagcgcacttaagattgcttaccgcattgaacggctggagtggcagactcgagagtgg 1949
Db    1681 TCCAGCGCACTTAAGATTGCTTACCGCATTGAACGGCTGGAGTGGCAGACTCGAGAGTGG 1740

Qy    1950 ccaatgcccgaggacctggctgaggctgatgacgagggtgagcctgaggagggaacttcc 2009
Db    1741 CCAATGCCCGAGGACCTGGCTGAGGCTGATGACGAGGGTGAGCCTGAGGAGGGAACTTCC 1800

Qy    2010 aaagatgaagagggcagggccagccaggagggaccccccaggtccagagtgaccccatccc 2069
Db    1801 AAAGATGAAGAGGGCAGGGCCAGCCAGGAGGGACCCCCAGGTCCAGAGTGACCCCATCCC 1860

Qy    2070 catccccaggattcccacctcctagtgaatccagagttgtagtaaagcctaactgctgca 2129
Db    1861 CATCCCCAGGATTCCCACCTCCTAGTGAATCCAGAGTTGTAGTAAAGCCTAACTGCTGCA 1920

Qy    2130 actctgtcatcctgtctgcgagatcacagacacaggagcgaattggtctgtagatgccca 2189
Db    1921 ACTCTGTCATCCTGTCTGCGAGATCACAGACACAGGAGCGAATTGGTCTGTAGATGCCCA 1980

Qy    2190 gctagagatataggagtttaacgcacattcagc 2222
Db    1981 GCTAGAGATATAGGAGTTTAACGCACATTCAGC 2013


RESULT   3
AX350402
LOCUS       AX350402                1835 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 27 from Patent WO0181578.
ACCESSION   AX350402
VERSION     AX350402.1  GI:18616045
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Vernet,C.A., Fernandes,E.R., Gerlach,V., Shimkets,R.A.,
            Malyankar,U.M., Boldog,F.L., Zerhusen,B.D., Spytek,K.A.,
            Majumder,K., Tchernev,V.T., Padigaru,M., Patturajan,M.,
            Burgess,C.E., Gangolli,E.A., Smithson,G., Rastelli,L.,
            Macdougall,J.R., Taupier,R.J., Grosse,W.M. and Alsobrook,J.P.
  TITLE     Novel proteins and nucleic acids encoding same
  JOURNAL   Patent: WO 0181578-A 27 01-NOV-2001;
            Curagen Corporation (US)
FEATURES             Location/Qualifiers
     source          1. .1835
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      420 a    521 c    521 g    373 t
ORIGIN


  Query Match             77.4%;  Score 1785.4;  DB 6;  Length 1835;
  Best Local Similarity   99.5%;  Pred. No. 0;
  Matches 1825;  Conservative    0;  Mismatches    1;  Indels    9;  Gaps    3;

Qy     280 tacaggcagagagggtgtggacatctcacaccccagcaccagaccacagaaccatgagcc 339
Db       1 TACAGGCAGAGAGGGTGTGGACATCTCACACCCCAGCACCAGACCACAGAACCATGAGCC 60

Qy     340 aggacaccaaagtgaagacaacagagtccagtcccccagccccatccaaggcc---agga 396
Db      61 AGGACACCAAAGTGAAGACAACAGAGTCCAGTCCCCCAGCCCCATCCAAGGCCAGGAGGA 120

Qy     397 agttgctgcctgtcctggacccatctggggattactactactggtggctgaacacaatgg 456
Db     121 AGTTGCTGCCTGTCCTGGACCCATCTGGGGATTACTACTACTGGTGGCTGAACACAATGG 180

Qy     457 tcttcccagtcatgtataacctcatcatcctcgtgtgcagagcctgcttccccgacttgc 516
Db     181 TCTTCCCAGTCATGTATAACCTCATCATCCTCGTGTGCAGGGCCTGCTTCCCCGACTTGC 240

Qy     517 agcacggttatctggtggcctggttggtgctggactacacgagtgacctgctatacctac 576
Db     241 AGCACGGTTATCTGGTGGCCTGGTTGGTGCTGGACTACACGAGTGACCTGCTATACCTAC 300

Qy     577 tagacatggtggtgcgcttccacaca---ggattcttggaacagggcatcctggtggtgg 633
Db     301 TAGACATGGTGGTGCGCTTCCACACAGGTGGATTCTTGGAACAGGGCATCCTGGTGGTGG 360

Qy     634 acaagggtaggatctcgagtcgctacgttcgcacctggagtttcttcttggacctggctt 693
Db     361 ACAAGGGTAGGATCTCGAGTCGCTACGTTCGCACCTGGAGTTTCTTCTTGGACCTGGCTT 420

Qy     694 ccctgatgcccacagatgtggtctacgtgcggctgggcccgcacacacccaccctgaggc 753
Db     421 CCCTGATGCCCACAGATGTGGTCTACGTGCGGCTGGGCCCGCACACACCCACCCTGAGGC 480

Qy     754 tgaaccgctttctccgcgcgccccgcctcttcgaggccttcgaccgcacagagacccgca 813
Db     481 TGAACCGCTTTCTCCGCGCGCCCCGCCTCTTCGAGGCCTTCGACCGCACAGAGACCCGCA 540

Qy     814 cagcttacccaaatgcctttcgcattgccaagctgatgctttacatttttgtcgtcatcc 873
Db     541 CAGCTTACCCAAATGCCTTTCGCATTGCCAAGCTGATGCTTTACATTTTTGTCGTCATCC 600

Qy     874 attggaacagctgcctatactttgccctatcccggtacctgggcttcgggcgtgacgcat 933
Db     601 ATTGGAACAGCTGCCTATACTTTGCCCTATCCCGGTACCTGGGCTTCGGGCGTGACGCAT 660

Qy     934 gggtgtacccggaccccgcgcagcctggctttgagcgcctgcggcgccagtacctctata 993
Db     661 GGGTGTACCCGGACCCCGCGCAGCCTGGCTTTGAGCGCCTGCGGCGCCAGTACCTCTATA 720

Qy     994 gcttttacttctccacgctgatactgactacagtgggcgatacaccgccgccagccaggg 1053
Db     721 GCTTTTACTTCTCCACGCTGATACTGACTACAGTGGGCGATACACCGCCGCCAGCCAGGG 780

Qy    1054 aagaagagtacctcttcatggtgggcgacttcctgctggccgtcatgggtttcgccacca 1113
Db     781 AAGAAGAGTACCTCTTCATGGTGGGCGACTTCCTGCTGGCCGTCATGGGTTTCGCCACCA 840

Qy    1114 tcatgggtagcatgagctctgtcatctacaacatgaacactgcagatgcggctttctacc 1173
Db     841 TCATGGGTAGCATGAGCTCTGTCATCTACAACATGAACACTGCAGATGCGGCTTTCTACC 900

Qy    1174 cagatcatgcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggagc 1233
Db     901 CAGATCATGCACTGGTGAAGAAGTACATGAAGCTGCAGCACGTCAACCGCAAGCTGGAGC 960

Qy    1234 ggcgagttattgactggtatcagcacctgcagatcaacaagaagatgaccaacgaggtag 1293
Db     961 GGCGAGTTATTGACTGGTATCAGCACCTGCAGATCAACAAGAAGATGACCAACGAGGTAG 1020

Qy    1294 ccatcttacagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctgt 1353
Db    1021 CCATCTTACAGCACTTGCCTGAGCGGCTGCGGGCAGAAGTGGCTGTGTCTGTGCACCTGT 1080

Qy    1354 ccactctgagccgggtgcagatctttcagaactgtgaggccagcctgctggaggagctgg 1413
Db    1081 CCACTCTGAGCCGGGTGCAGATCTTTCAGAACTGTGAGGCCAGCCTGCTGGAGGAGCTGG 1140

Qy    1414 tgctgaagctgcagccccagacctactcaccaggtgaatatgtatgccgcaaaggagaca 1473
Db    1141 TGCTGAAGCTGCAGCCCCAGACCTACTCACCAGGTGAATATGTATGCCGCAAAGGAGACA 1200

Qy    1474 ttggccaagagatgtacatcatccgagagggtcaactggccgtggtggcagatgatggta 1533
Db    1201 TTGGCCAAGAGATGTACATCATCCGAGAGGGTCAACTGGCCGTGGTGGCAGATGATGGTA 1260

Qy    1534 tcacacagtatgctgtgctcggtgcagggctctactttggggagatcagcatcatcaaca 1593
```

```
Db   1261 TCACACAGTATGCTGTGCTCGGTGCAGGGCTCTACTTTGGGGAGATCAGCATCATCAACA 1320

Qy   1594 tcaaa---gggaacatgtctgggaaccgccgcacagccaacatcaagagcctaggttatt 1650
          |||||   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 TCAAAGGTGGGAACATGTCTGGGAACCGCCGCACAGCCAACATCAAGAGCCTAGGTTATT 1380

Qy   1651 cagacctattctgcctgagcaaggaggacctgcgggaggtgctgagcgagtatccacaag 1710
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 CAGACCTATTCTGCCTGAGCAAGGAGGACCTGCGGGAGGTGCTGAGCGAGTATCCACAAG 1440

Qy   1711 cacagaccatcatggaggagaaaggacgtgagatcctgctgaaaatgaacaagttggacg 1770
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 CACAGACCATCATGGAGGAGAAAGGACGTGAGATCCTGCTGAAAATGAACAAGTTGGACG 1500

Qy   1771 tgaatgctgaggcagctgagatcgccctgcaggaggccacagagtcccggctacgaggcc 1830
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 TGAATGCTGAGGCAGCTGAGATCGCCCTGCAGGAGGCCACAGAGTCCCGGCTACGAGGCC 1560

Qy   1831 tagaccagcagctggatgatctacagaccaagtttgctcgcctcctggctgagctggagt 1890
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 TAGACCAGCAGCTGGATGATCTACAGACCAAGTTTGCTCGCCTCCTGGCTGAGCTGGAGT 1620

Qy   1891 ccagcgcacttaagattgcttaccgcattgaacggctggagtggcagactcgagagtggc 1950
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 CCAGCGCACTTAAGATTGCTTACCGCATTGAACGGCTGGAGTGGCAGACTCGAGAGTGGC 1680

Qy   1951 caatgcccgaggacctggctgaggctgatgacgagggtgagcctgaggagggaacttcca 2010
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 CAATGCCCGAGGACCTGGCTGAGGCTGATGACGAGGGTGAGCCTGAGGAGGGAACTTCCA 1740

Qy   2011 aagatgaagagggcagggccagccaggagggaccccccaggtccagagtgaccccatcccc 2070
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 AAGATGAAGAGGGCAGGGCCAGCCAGGAGGGACCCCCAGGTCCAGAGTGACCCCATCCCC 1800

Qy   2071 atccccaggattcccacctcctagtgaatccagag 2105
          |||||||||||||||||||||||||||||||||
Db   1801 ATCCCCAGGATTCCCACCTCCTAGTGAATCCAGAG 1835


RESULT   4
RNU12623
LOCUS       RNU12623                2351 bp    mRNA    linear   ROD 15-SEP-1994
DEFINITION  Rattus norvegicus Sprague-Dawley cyclic nucleotide gated cation
            channel mRNA, complete cds.
ACCESSION   U12623
VERSION     U12623.1  GI:538128
KEYWORDS    .
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 2351)
  AUTHORS   Bradley,J.; Li,J., Davidson,N., Lester,H.A. and Zinn,K.
  TITLE     Heteromeric olfactory cyclic nucleotide-gated channels: A new
            subunit that confers increased sensitivity to cAMP
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 91, 8890-8894 (1994)
  MEDLINE   94377458
REFERENCE   2  (bases 1 to 2351)
  AUTHORS   Bradley,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-1994) Jonathan Bradley, Biology, California
            Institute of Technology, Pasadena, CA 91125, USA
COMMENT     On Sep 15, 1994 this sequence version replaced gi:521147.
FEATURES             Location/Qualifiers
     source          1. .2351
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
     CDS             109. .1836
                     /codon_start=1
                     /product="cyclic nucleotide gated cation channel"
                     /protein_id="AAA21464.1"
                     /db_xref="GI:538129"
                     /translation="MSQDGKVKTTESTPPAPTKARKWLPVLDPSGDYYYWWLNTMVFP
                     IMYNLIIVVCRACFPDLQHSYLVAWFVLDYTSDLLYLLDIGVRFHTGFLEQGILVVDK
                     GMIASRYVRTWSFLLDLASLVPTDAAYVQLGPHIPTLRLNRFLRVPRLFEAFDRTETR
                     TAYPNAFRIAKLMLYIFVVIHWNSCLYFALSRYLGFGRDAWVYPDPAQPGFERLRRQY
                     LYSFYFSTLILTTVGDTPLPDREEEYLFMVGDFLLAVMGFATIMGSMSSVIYNMNTAD
                     AAFYPDHALVKKYMKLQHVNKRLERRVIDWYQHLQINKKMTNEVAILQHLPERLRAEV
                     AVSVHLSTLSRVQIFQNCEASLLEELVLKLQPQTYSPGEYVCRKGDIGREMYIIREGQ
                     LAVVADDGVTQYAVLGAGLYFGEISIINIKGNMSGNRRTANIKSLGYSDLFCLSKEDL
                     REVLSEYPQAQAVMEEKGREILLKMNKLDVNAEAAEIALQEATESRLKGLDQQLDDLQ
                     TKFARLLAELESSALKIAYRIERLEWQTREWPMPEDMGEADDEAEPGEGTSKDGEGKA
                     GQAGPSGIE"
BASE COUNT      580 a    624 c    637 g    510 t
ORIGIN


  Query Match           65.4%;  Score 1508.4;  DB 10;  Length 2351;
  Best Local Similarity  85.1%;  Pred. No. 0;
  Matches 1759;  Conservative    0;  Mismatches  291;  Indels   16;  Gaps    6;

Qy    216 gcagaagcccaacagcagcctccttcagacagtcaggcactagtgcccaactccagaagt 275
          || | ||||| |  |||||||| | |||  || ||||||||          | |  |||
Db      2 GCTGGAGCCCGAGGGCAGCCTCTTCCAGGAAGCCAGGCACT----------TTCCCAAGC 51

Qy    276 cccctacaggcagagagggtgtggacatctcacaccccagcaccagaccacagaaccatg 335
          | |||| |  | ||||     |||| |    |||  | || |||| || || ||| ||||
Db     52 CACCTATAATCGGAGAATCACTGGATAGTCTACAGACTAGAACCACACTACGGAAACATG 111

Qy    336 agccaggacaccaaagtgaagacaacagagtccagtcccccagccccatccaaggccagg 395
          |||||||||  |||||||||||| ||||||||||  |||||||||||| |||| ||||||
Db    112 AGCCAGGACGGCAAAGTGAAGACCACAGAGTCCACGCCCCCAGCCCCAACCAAAGCCAGG 171

Qy    396 aagttgctgcctgtcctggacccatctggggattactactactggtggctgaacacaatg 455
          |||| |||||| ||||| ||||||||||||||||||||||||||||||||||||||||||
Db    172 AAGTGGCTGCCAGTCCTAGACCCATCTGGGGATTACTACTACTGGTGGCTGAACACAATG 231

Qy    456 gtcttcccagtcatgtataacctcatcatcctcgtgtgcagagcctgcttccccgacttg 515
          ||||||||| ||||||| |||||||||||| | || ||||| |||||||| || ||||||
Db    232 GTCTTCCCAATCATGTACAACCTCATCATCGTTGTATGCAGGGCCTGCTTTCCTGACTTG 291

Qy    516 cagcacggttatctggtggcctggttggtgctggactacacgagtgacctgctataccta 575
          |||||| |||| |||||||||||||| ||||||||||||||||||||||||||| ||||||
Db    292 CAGCACAGTTACCTGGTGGCCTGGTTCGTGCTGGACTACACGAGTGACCTGCTGTACCTA 351

Qy    576 ctagacatggtggtgcgcttccacacaggattcttggaacagggcatcctggtggtggac 635
          ||||| || | ||| |||||||||||||||||| | || |||||||||||||||||| |||
Db    352 CTAGATATTGGGGTACGCTTCCACACAGGATTCCTAGAGCAGGGCATCCTGGTGGTAGAC 411

Qy    636 aagggtaggatctcgagtcgctacgttcgcacctggagtttcttcttggacctggcttcc 695
          || || | |||| | |||||||||| |||||||||| ||| | ||||||||||| |||
Db    412 AAAGGCATGATCGCCAGTCGCTACGTCCGCACCTGGAGCTTCCTGTTGGACCTGGCGTCC 471

Qy    696 ctgatgcccacagatgtggtctacgtgcggctgggcccgcacacacccaccctgaggctg 755
          ||| | |||||||||| || ||| |||| |||||||| ||||  || || ||  ||||
Db    472 CTGGTCCCCACAGATGCGGCCTATGTGCAGCTGGGCCCCCACATCCCTACACTCCGGCTA 531

Qy    756 aaccgctttctccgcgcgccccgcctcttcgaggccttcgaccgcacagagacccgcaca 815
          |||||||||||||| | |||||||||||||||||| || || || ||||||||||||||
Db    532 AACCGCTTTCTCCGAGTGCCCCGCCTCTTCGAGGCTTTTGATCGTACAGAGACCCGCACG 591

Qy    816 gcttacccaaatgcctttcgcattgccaagctgatgctttacatttttgtcgtcatccat 875
          ||||||||||||||||| |||| |||||||||||||||||||||||||| |||||||||
Db    592 GCTTACCCAAATGCCTTCCGCATAGCCAAGCTGATGCTTTACATTTTTGTTGTCATCCAT 651

Qy    876 tggaacagctgcctatactttgccctatcccggtacctgggcttcgggcgtgacgcatgg 935
          |||||||| ||| ||||||| ||||| ||| |||||||||||||| || || || || |||
Db    652 TGGAACAGTTGCTTATACTTCGCCCTGTCCAGGTACCTGGGCTTTGGACGGGATGCGTGG 711

Qy    936 gtgtacccggaccccgcgcagcctggctttgagcgcctgcggcgccagtacctctatagc 995
          || ||||| ||||| ||||| |||||||||||||| |||||||||||| ||||| |||
Db    712 GTATACCCAGACCCTGCGCAACCTGGCTTTGAGCGCTTGCGGCGCCAGTATCTCTACAGC 771

Qy    996 ttttacttctccacgctgatactgactacagtgggcgatacaccgccgccagccagggaa 1055
```

```
Db    772 TTCTACTTCTCCACTCTGATCCTGACCACAGTGGGTGACACGCCGCTGCCAGACCGAGAG 831

Qy   1056 gaagagtacctcttcatggtgggcgacttcctgctggccgtcatgggtttcgccaccatc 1115
Db    832 GAAGAGTACCTCTTCATGGTGGGTGACTTCCTGCTGGCCGTCATGGGTTTCGCCACCATC 891

Qy   1116 atgggtagcatgagctctgtcatctacaacatgaacactgcagatgcggctttctaccca 1175
Db    892 ATGGGTAGCATGAGCTCTGTCATTTACAACATGAACACTGCAGATGCGGCCTTCTACCCA 951

Qy   1176 gatcatgcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggagcgg 1235
Db    952 GACCATGCGCTGGTAAAGAAGTACATGAAGCTGCAGCATGTCAACAAGAGGCTGGAGCGG 1011

Qy   1236 cgagttattgactggtatcagcacctgcagatcaacaagaagatgaccaacgaggtagcc 1295
Db   1012 CGAGTTATTGACTGGTACCAGCATCTTCAGATCAACAAGAAGATGACCAACGAGGTAGCC 1071

Qy   1296 atcttacagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctgtcc 1355
Db   1072 ATCTTGCAGCACCTGCCTGAGCGGCTGCGGGCGGAGGTTGCTGTGTCCGTGCACCTGTCT 1131

Qy   1356 actctgagccgggtgcagatctttcagaactgtgaggccagcctgctggaggagctggtg 1415
Db   1132 ACCCTGAGCCGAGTACAGATCTTCCAGAACTGTGAAGCCAGCCTGCTGGAAGAGCTGGTG 1191

Qy   1416 ctgaagctgcagccccagacctactcaccaggtgaatatgtatgccgcaaaggagacatt 1475
Db   1192 CTGAAGCTACAGCCCCAGACCTACTCGCCAGGCGAATATGTGTGCCGCAAAGGGGACATT 1251

Qy   1476 ggccaagagatgtacatcatccgagagggtcaactggccgtggtggcagatgatggtatc 1535
Db   1252 GGCCGAGAGATGTACATCATCCGTGAGGGCCAGCTGGCTGTGGTGGCCGATGATGGTGTC 1311

Qy   1536 acacagtatgctgtgctcggtgcagggctctactttggggagatcagcatcatcaacatc 1595
Db   1312 ACACAGTATGCTGTGCTTGGTGCAGGGCTCTACTTTGGGGAGATCAGTATCATCAACATC 1371

Qy   1596 aaagggaacatgtctgggaaccgccgcacagccaacatcaagagcctaggttattcagac 1655
Db   1372 AAAGGGAACATGTCTGGAAACCGACGAACAGCCAACATCAAGAGCCTAGGTTATTCAGAC 1431

Qy   1656 ctattctgcctgagcaaggaggacctgcgggaggtgctgagcgagtatccacaagcacag 1715
Db   1432 CTGTTCTGCCTCAGCAAGGAGGATCTGCGGGAGGTACTGAGTGAGTACCCACAGGCCCAG 1491

Qy   1716 accatcatggaggagaaaggacgtgagatcctgctgaaaatgaacaagttggacgtgaat 1775
Db   1492 GCGGTCATGGAGGAGAAGGGCCGAGAAATCTTGCTCAAAATGAATAAGTTGGATGTGAAT 1551

Qy   1776 gctgaggcagctgagatcgccctgcaggaggccacagagtcccggctacgaggcctagac 1835
Db   1552 GCTGAGGCAGCTGAGATCGCCCTCCAGGAGGCCACAGAGTCTCGGCTCAAAGGCCTCGAC 1611

Qy   1836 cagcagctggatgatctacagaccaagtttgctcgcctcctggctgagctggagtccagc 1895
Db   1612 CAGCAGCTTGATGATCTGCAGACCAAGTTTGCTCGCCTACTGGCTGAGCTGGAGTCCAGT 1671

Qy   1896 gcacttaagattgcttaccgcattgaacggctggagtggcagactcgagagtggccaatg 1955
Db   1672 GCACTGAAGATAGCTTACCGCATCGAAAGGCTGGAGTGGCAGACTCGAGAGTGGCCAATG 1731

Qy   1956 cccgaggacctggctgaggctgatgacgagggtgagcctgaggagggaacttccaaagat 2015
Db   1732 CCAGAGGACATGGGTGAGGCTGATGATGAGGCTGAGCCTGGAGAAGGGACGTCCAAGGAT 1791

Qy   2016 gaagagggcagggccagccaggagggacccccaggtccagagtgaccccatccccatccc 2075
Db   1792 GGAGAGGGAAAGGCTGGCCAGGCGGGACCCTCAGGCATAGAATGACCCCGTCCTGA-CCC 1850

Qy   2076 caggattcccacctcctagtgaatccagagttgtagtaaagcctaactgctgcaactctg 2135
```

```
Db   1851 TAGGACTCCCAGCTCAAA-TAAATCCAGAGTGGTGGGAAAGTCTGCCTGCAGGAA--CTG 1907

Qy   2136 tcatcctgtctgcgagatcacagacacaggagcgaattggtctgtagatgcccagctaga 2195
Db   1908 TCATCCTCTTTGCTAGGTTACAGA-ACGTAGGTAAATTGGTCTATAGATGCCTAGCTAGG 1966

Qy   2196 gatataggagtttaacgcacattcagccccccacttaccagtacacacacacacacacaca 2255
Db   1967 GATGT-GGGTCACAGCATCCATTAGTCCCATACTCACCAGCAAATGTACACACACACACA 2025

Qy   2256 cacacacatttgctcatagacctgtt 2281
Db   2026 CACACAAACATGCGCACGCACACGCT 2051


RESULT   5
RNU12425
LOCUS       RNU12425                1938 bp    mRNA    linear   ROD 05-APR-1995
DEFINITION  Rattus norvegicus olfactory cyclic nucleotide-gated channel mRNA,
            complete cds.
ACCESSION   U12425
VERSION     U12425.1  GI:548083
KEYWORDS    .
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1938)
  AUTHORS   Liman,E.R. and Buck,L.B.
  TITLE     A second subunit of the olfactory cyclic nucleotide-gated channel
            confers high sensitivity to cAMP
  JOURNAL   Neuron 13 (3), 611-621 (1994)
  MEDLINE   95000663
REFERENCE   2  (bases 1 to 1938)
  AUTHORS   Liman,E.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-1994) Emily R. Liman, Neurobiology, Harvard
            Medical School, 220 Longwood Avenue, Boston, MA 02115, USA
FEATURES             Location/Qualifiers
     source          1. .1938
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /tissue_type="olfactory neuroepithelium"
     mRNA            1. .1937
                     /evidence=experimental
     CDS             107. .1834
                     /codon_start=1
                     /product="olfactory cyclic nucleotide-gated channel"
                     /protein_id="AAA64748.1"
                     /db_xref="GI:548084"
                     /translation="MSQDGKVKTTESTPPAPTKARKWLPVLDPSGDYYYWWLNTMVFP
                     IMYNLIIVVCRACFPDLQHSYLVAWFVLDYTSDLLYLLDIGVRFHTGFLEQGILVVDK
                     GMIASRYVRTWSFLLDLASLVPTDAAYVQLGPHIPTLRLNRFLRVPRLFEAFDRTETR
                     TAYPNAFRIAKLMLYIFVVIHWNSCLYFALSRYLGFGRDAWVYPDPAQPGFERLRRQY
                     LYSFYFSTLILTTVGDTPLPDREEEYLFMVGDFLLAVMGFATIMGSMSSVIYNMNTAD
                     AAFYPDHALVKKYMKLQHVNKRLERRVIDWYQHLQINKKMTNEVAILQHLPERLRAEV
                     AVSVHLSTLSRVQIFQNCEASLLEELVLKLQPQTYSPGEYVCRKGDIGREMYIIREGQ
                     LAVVADDGVTQYAVLGAGLYFGEISIINIKGNMSGNRRTANIKSLGYSDLFCLSKEDL
                     REVLSEYPQAQAVMEEKGREILLKMNKLDVNAEAAEIALQEATESRLKGLDQQLDDLQ
                     TKFARLLAELESSALKIAYRIERLEWQTREWPMPEDMGEADDEAEPGEGTSKDGEGKA
                     GQAGPSGIE"
     polyA_site      1938
                     /note="3 A residues"
BASE COUNT      458 a    531 c    543 g    406 t
ORIGIN


  Query Match           63.7%;  Score 1470.2;  DB 10;  Length 1938;
  Best Local Similarity   87.0%;  Pred. No. 0;
  Matches 1650;  Conservative    0;  Mismatches  243;  Indels    4;  Gaps    3;

Qy    263 caactccagaagtcccctacaggcagagagggtgtggacatctcacaccccagcaccaga 322
```

```
Db      37 CACTTTCCCAAGCCACCTATAATCGGAGAATCACTGGATAGTCTACAGACTAGAACCACA 96

Qy     323 ccacagaaccatgagccaggacaccaaagtgaagacaacagagtccagtcccccagcccc 382
Db      97 CTACGGAAACATGAGCCAGGACGGCAAAGTGAAGACCACAGAGTCCACGCCCCCAGCCCC 156

Qy     383 atccaaggccaggaagttgctgcctgtcctggacccatctggggattactactactggtg 442
Db     157 AACCAAAGCCAGGAAGTGGCTGCCAGTCCTAGACCCATCTGGGGATTACTACTACTGGTG 216

Qy     443 gctgaacacaatggtcttcccagtcatgtataacctcatcatcctcgtgtgcagagcctg 502
Db     217 GCTGAACACAATGGTCTTCCCAATCATGTACAACCTCATCATCGTTGTATGCAGGGCCTG 276

Qy     503 cttccccgacttgcagcacggttatctggtggcctggttggtgctggactacacgagtga 562
Db     277 CTTTCCTGACTTGCAGCACAGTTACCTGGTGGCCTGGTTCGTGCTGGACTACACGAGTGA 336

Qy     563 cctgctatacctactagacatggtggtgcgcttccacacaggattcttggaacagggcat 622
Db     337 CCTGCTGTACCTACTAGATATTGGGGTACGCTTCCACACAGGATTCCTAGAGCAGGGCAT 396

Qy     623 cctggtggtggacaagggtaggatctcgagtcgctacgttcgcacctggagtttcttctt 682
Db     397 CCTGGTGGTAGACAAAGGCATGATCGCCAGTCGCTACGTCCGCACCTGGAGCTTCCTGTT 456

Qy     683 ggacctggcttccctgatgcccacagatgtggtctacgtgcggctgggcccgcacacacc 742
Db     457 GGACCTGGCGTCCCTGGTCCCCACAGATGCGGCCTATGTGCAGCTGGGCCCCCACATCCC 516

Qy     743 caccctgaggctgaaccgctttctccgcgcgccccgcctcttcgaggccttcgaccgcac 802
Db     517 TACACTCCGGCTAAACCGCTTTCTCCGAGTGCCCCGCCTCTTCGAGGCTTTTGATCGTAC 576

Qy     803 agagacccgcacagcttacccaaatgcctttcgcattgccaagctgatgctttacatttt 862
Db     577 AGAGACCCGCACGGCTTACCCAAATGCCTTCCGCATAGCCAAGCTGATGCTTTACATTTT 636

Qy     863 tgtcgtcatccattggaacagctgcctatactttgccctatcccggtacctgggcttcgg 922
Db     637 TGTTGTCATCCATTGGAACAGTTGCTTATACTTCGCCCTGTCCAGGTACCTGGGCTTTGG 696

Qy     923 gcgtgacgcatgggtgtacccggaccccgcgcagcctggctttgagcgcctgcggcgcca 982
Db     697 ACGGGATGCGTGGGTATACCCAGACCCTGCGCAACCTGGCTTTGAGCGCTTGCGGCGCCA 756

Qy     983 gtacctctatagcttttacttctccacgctgatactgactacagtgggcgatacaccgcc 1042
Db     757 GTATCTCTACAGCTTCTACTTCTCCACTCTGATCCTGACCACAGTGGGTGACACGCCGCT 816

Qy    1043 gccagccagggaagaagagtacctcttcatggtgggcgacttcctgctggccgtcatggg 1102
Db     817 GCCAGACCGAGAGGAAGAGTACCTCTTCATGGTGGGTGACTTCCTGCTGGCCGTCATGGG 876

Qy    1103 tttcgccaccatcatgggtagcatgagctctgtcatctacaacatgaacactgcagatgc 1162
Db     877 TTTCGCCACCATCATGGGTAGCATGAGCTCTGTCATTTACAACATGAACACTGCAGATGC 936

Qy    1163 ggctttctacccagatcatgcactggtgaagaagtacatgaagctgcagcacgtcaaccg 1222
Db     937 GGCCTTCTACCCAGACCATGCGCTGGTAAAGAAGTACATGAAGCTGCAGCATGTCAACAA 996

Qy    1223 caagctggagcggcgagttattgactggtatcagcacctgcagatcaacaagaagatgac 1282
Db     997 GAGGCTGGAGCGGCGAGTTATTGACTGGTACCAGCATCTTCAGATCAACAAGAAGATGAC 1056

Qy    1283 caacgaggtagccatcttacagcacttgcctgagcggctgcgggcagaagtggctgtgtc 1342
Db    1057 CAACGAGGTAGCCATCTTGCAGCACCTGCCTGAGCGGCTGCGGGCGGAGGTTGCTGTGTC 1116

Qy    1343 tgtgcacctgtccactctgagccgggtgcagatctttcagaactgtgaggccagcctgct 1402
Db    1117 CGTGCACCTGTCTACCCTGAGCCGAGTACAGATCTTCCAGAACTGTGAAGCCAGCCTGCT 1176

Qy    1403 ggaggagctggtgctgaagctgcagccccagacctactcaccaggtgaatatgtatgccg 1462
Db    1177 GGAAGAGCTGGTGCTGAAGCTACAGCCCCAGACCTACTCGCCAGGCGAATATGTGTGCCG 1236

Qy    1463 caaaggagacattggccaagagatgtacatcatccgagagggtcaactggccgtggtggc 1522
Db    1237 CAAAGGGGACATTGGCCGAGAGATGTACATCATCCGTGAGGGCCAGCTGGCTGTGGTGGC 1296

Qy    1523 agatgatggtatcacacagtatgctgtgctcggtgcagggctctactttggggagatcag 1582
Db    1297 CGATGATGGTGTCACACAGTATGCTGTGCTTGGTGCAGGGCTCTACTTTGGGGAGATCAG 1356

Qy    1583 catcatcaacatcaaagggaacatgtctgggaaccgccgcacagccaacatcaagagcct 1642
Db    1357 TATCATCAACATCAAAGGGAACATGTCTGGAAACCGACGAACAGCCAACATCAAGAGCCT 1416

Qy    1643 aggttattcagacctattctgcctgagcaaggaggacctgcgggaggtgctgagcgagta 1702
Db    1417 AGGTTATTCAGACCTGTTCTGCCTCAGCAAGGAGGATCTGCGGGAGGTACTGAGTGAGTA 1476

Qy    1703 tccacaagcacagaccatcatggaggagaaaggacgtgagatcctgctgaaaatgaacaa 1762
Db    1477 CCCACAGGCCCAGGCGGTCATGGAGGAGAAGGGCCGAGAAATCTTGCTCAAAATGAATAA 1536

Qy    1763 gttggacgtgaatgctgaggcagctgagatcgccctgcaggaggccacagagtcccggct 1822
Db    1537 GTTGGATGTGAATGCTGAGGCAGCTGAGATCGCCCTCCAGGAGGCCACAGAGTCTCGGCT 1596

Qy    1823 acgaggcctagaccagcagctggatgatctacagaccaagtttgctcgcctcctggctga 1882
Db    1597 CAAAGGCCTCGACCAGCAGCTTGATGATCTGCAGACCAAGTTTGCTCGCCTACTGGCTGA 1656

Qy    1883 gctggagtccagcgcacttaagattgcttaccgcattgaacggctggagtggcagactcg 1942
Db    1657 GCTGGAGTCCAGTGCACTGAAGATAGCTTACCGCATCGAAAGGCTGGAGTGGCAGACTCG 1716

Qy    1943 agagtggccaatgcccgaggacctggctgaggctgatgacgagggtgagcctgaggaggg 2002
Db    1717 AGAGTGGCCAATGCCAGAGGACATGGGTGAGGCTGATGATGAGGCTGAGCCTGGAGAAGG 1776

Qy    2003 aacttccaaagatgaagagggcagggccagccaggagggacccccaggtccagagtgacc 2062
Db    1777 GACGTCCAAGGATGGAGAGGGAAAGGCTGGCCAGGCGGGACCCTCAGGCATAGAATGACC 1836

Qy    2063 ccatccccatccccaggattcccacctcctagtgaatccagagttgtagtaaagcctaac 2122
Db    1837 CCGTCCTGA-CCCTAGGACTCCCAGCT-CAAATAAATCCAGAGTGGTGGGAAAGTCTGCC 1894

Qy    2123 tgctgcaactctgtcatcctgtctgcgagatcacaga 2159
Db    1895 TGCAGGAA--CTGTCATCCTCTTTGCTAGGTTACAGA 1929
```

```
RESULT   6
AC036216
LOCUS       AC036216              164434 bp    DNA     linear   HTG 04-MAY-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-658A5 map 11, WORKING DRAFT
            SEQUENCE, 26 unordered pieces.
ACCESSION   AC036216
VERSION     AC036216.2  GI:7705203
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 164434)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-658A5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 164434)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
```

```
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 4, 2000 this sequence version replaced gi:7523886.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L9404
                Center clone name: 658_A_5
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 149016 bases at least Q40
                Consensus quality: 157301 bases at least Q30
                Consensus quality: 160298 bases at least Q20
                Insert size: 157000; agarose-fp
                Insert size: 161934; sum-of-contigs
                Quality coverage: 3.7 in Q20 bases; agarose-fp
                Quality coverage: 3.6 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1416: contig of 1416 bp in length
            *     1417 1516: gap of      100 bp
            *     1517     3580: contig of 2064 bp in length
            *     3581 3680: gap of      100 bp
            *     3681     4383: contig of 703 bp in length
            *     4384 4483: gap of      100 bp
            *     4484     5844: contig of 1361 bp in length
            *     5845 5944: gap of      100 bp
            *     5945     7574: contig of 1630 bp in length
            *     7575 7674: gap of      100 bp
            *     7675     9988: contig of 2314 bp in length
            *     9989 10088: gap of      100 bp
            *    10089    13426: contig of 3338 bp in length
            *    13427 13526: gap of      100 bp
            *    13527    15573: contig of 2047 bp in length
            *    15574 15673: gap of      100 bp
            *    15674    18974: contig of 3301 bp in length
            *    18975 19074: gap of      100 bp
            *    19075    21809: contig of 2735 bp in length
            *    21810 21909: gap of      100 bp
            *    21910    25085: contig of 3176 bp in length
            *    25086 25185: gap of      100 bp
            *    25186    29425: contig of 4240 bp in length
            *    29426 29525: gap of      100 bp
            *    29526    34133: contig of 4608 bp in length
            *    34134 34233: gap of      100 bp
            *    34234    40649: contig of 6416 bp in length
            *    40650 40749: gap of      100 bp
            *    40750    45891: contig of 5142 bp in length
            *    45892 45991: gap of      100 bp
            *    45992    53180: contig of 7189 bp in length
            *    53181 53280: gap of      100 bp
            *    53281    62817: contig of 9537 bp in length
            *    62818 62917: gap of      100 bp
            *    62918    70247: contig of 7330 bp in length
            *    70248 70347: gap of      100 bp
            *    70348    78754: contig of 8407 bp in length
            *    78755 78854: gap of      100 bp
            *    78855    86840: contig of 7986 bp in length
            *    86841 86940: gap of      100 bp
            *    86941    98610: contig of 11670 bp in length
            *    98611 98710: gap of      100 bp
            *    98711   111069: contig of 12359 bp in length
            *   111070 111169: gap of      100 bp
            *   111170   122730: contig of 11561 bp in length
            *   122731 122830: gap of      100 bp
            *   122831   135722: contig of 12892 bp in length
            *   135723 135822: gap of      100 bp
            *   135823   150963: contig of 15141 bp in length
            *   150964 151063: gap of      100 bp
            *   151064   164434: contig of 13371 bp in length.
FEATURES             Location/Qualifiers
     source          1. .164434
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-658A5"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1416
                     /note="assembly_fragment"
     misc_feature    1517. .3580
                     /note="assembly_fragment"
     misc_feature    3681. .4383
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    4484. .5844
                     /note="assembly_fragment"
     misc_feature    5945. .7574
                     /note="assembly_fragment"
     misc_feature    7675. .9988
                     /note="assembly_fragment"
     misc_feature    10089. .13426
                     /note="assembly_fragment"
     misc_feature    13527. .15573
                     /note="assembly_fragment"
     misc_feature    15674. .18974
                     /note="assembly_fragment"
     misc_feature    19075. .21809
                     /note="assembly_fragment"
     misc_feature    21910. .25085
                     /note="assembly_fragment"
     misc_feature    25186. .29425
                     /note="assembly_fragment"
     misc_feature    29526. .34133
                     /note="assembly_fragment"
     misc_feature    34234. .40649
                     /note="assembly_fragment"
     misc_feature    40750. .45891
                     /note="assembly_fragment"
```

```
     misc_feature    45992. .53180
                     /note="assembly_fragment"
     misc_feature    53281. .62817
                     /note="assembly_fragment"
     misc_feature    62918. .70247
                     /note="assembly_fragment"
     misc_feature    70348. .78754
                     /note="assembly_fragment"
     misc_feature    78855. .86840
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    86941. .98610
                     /note="assembly_fragment"
     misc_feature    98711. .111069
                     /note="assembly_fragment"
     misc_feature    111170. .122730
                     /note="assembly_fragment"
     misc_feature    122831. .135722
                     /note="assembly_fragment"
     misc_feature    135823. .150963
                     /note="assembly_fragment"
     misc_feature    151064. .164434
                     /note="assembly_fragment"
BASE COUNT    50098 a  34031 c  34547 g  43240 t   2518 others
ORIGIN


  Query Match `          29.8%;  Score 687.2;  DB 2;  Length 164434;
  Best Local Similarity   97.9%;  Pred. No. 8.4e-155;
  Matches  711;  Conservative    0;  Mismatches    3;  Indels   12;  Gaps    1;

Qy   1595 caaagggaacatgtctgggaaccgccgcacagccaacatcaagagcctaggttattcaga 1654
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 161377 CACAGGGAACATGTCTGGGAACCGCCGCACAGCCAACATCAAGAGCCTAGGTTATTCAGA 161436

Qy   1655 cctattctgcctgagcaaggaggacctgcgggaggtgctgagcgagtatccacaagcaca 1714
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 161437 CCTATTCTGCCTGAGCAAGGAGGACCTGCGGGAGGTGCTGAGCGAGTATCCACAAGCACA 161496

Qy   1715 gaccatcatggaggagaaaggacgtgagatcctgctgaaaatgaacaagttggacgtgaa 1774
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 161497 GACCATCATGGAGGAGAAAGGACGTGAGATCCTGCTGAAAATGAACAAGTTGGACGTGAA 161556

Qy   1775 tgctgaggcagctgagatcgccctgcaggaggccacagagtcccggctacgaggcctaga 1834
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db 161557 TGCTGAGGCAGCTGAGATCGCCCTGCAGGAGGCCACAGAGTCCCGGCTACGAGGCATAGA 161616

Qy   1835 ccagcagctggatgatctacagaccaagtttgctcgcctcctggctgagctggagtccag 1894
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 161617 CCAGCAGCTGGATGATCTACAGACCAAGTTTGCTCGCCTCCTGGCTGAGCTGGAGTCCAG 161676

Qy   1895 cgcacttaagattgcttaccgcattgaacggctggagtggcagactcgagagtggccaat 1954
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 161677 CGCACTTAAGATTGCTTACCGCATTGAACGGCTGGAGTGGCAGACTCGAGAGTGGCCAAT 161736

Qy   1955 gcccgaggacctggctgaggctgatgacgagggtgagcctgaggagggaacttccaaaga 2014
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db 161737 GCCCGAGGACCTGGCTGAGGCTGATGACGAGGGTGTGCCTGAGGAGGGAACTTCCAAAGA 161796

Qy   2015 tgaagagggcagggccagccaggagggaccccccaggtccagagtgaccccatccccatcc 2074
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 161797 TGAAGAGGGCAGGGCCAGCCAGGAGGGACCCCCAGGTCCAGAGTGACCCCATCCCCATCC 161856

Qy   2075 ccaggattcccacctcctagtgaatccagagttgtagtaaagcctaactgctgcaactct 2134
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 161857 CCAGGATTCCCACCTCCTAGTGAATCCAGAGTTGTAGTAAAGCCTAACTGCTGCAACTCT 161916

Qy   2135 gtcatcctgtctgcgagatcacagacacaggagcgaattggtctgtagatgcccagctag 2194
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 161917 GTCATCCTGTCTGCGAGATCACAGACACAGGAGCGAATTGGTCTGTAGATGCCCAGCTAG 161976

Qy   2195 agatataggagtttaacgcacattcagcccccacttaccagt----------acacac 2242
          |||||||||||||||||||||||||||||||||||||||||          ||||||
Db 161977 AGATATAGGAGTTTAACGCACATTCAGCCCCCACTTACCAGTACACACACACACACACAC 162036

Qy   2243 acacacacacacacacacacatttgctcatagacctgttggccccaagactgtgcattcc 2302
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 162037 ACACACACACACACACACACATTTGCTCATAGACCTGTTGGCCCCAAGACTGTGCATTCC 162096

Qy   2303 atctaa 2308
          ||||||
Db 162097 ATCTAA 162102


RESULT   7
AC021935
LOCUS       AC021935              162965 bp    DNA     linear   HTG 03-MAR-2001
DEFINITION  Homo sapiens chromosome 11 clone RP11-645I8 map 11, WORKING DRAFT
            SEQUENCE, 5 unordered pieces.
ACCESSION   AC021935
VERSION     AC021935.3  GI:13184160
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 162965)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-645I8
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 162965)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 2, 2001 this sequence version replaced gi:7652890.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5277
                Center clone name: 645_I_8
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 40% of reads
                Sequencing vector: Plasmid; n/a; 60% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 161664 bases at least Q40
                Consensus quality: 161989 bases at least Q30
                Consensus quality: 162229 bases at least Q20
                Insert size: 158000; agarose-fp
                Insert size: 162565; sum-of-contigs
                Quality coverage: 12.2 in Q20 bases; agarose-fp
                Quality coverage: 11.9 in Q20.
```

```
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     5772: contig of 5772 bp in length
            *    5773 5872: gap of      100 bp
            *    5873   114225: contig of 108353 bp in length
            *  114226 114325: gap of      100 bp
            *  114326   119496: contig of 5171 bp in length
            *  119497 119596: gap of      100 bp
            *  119597   161781: contig of 42185 bp in length
            *  161782 161881: gap of      100 bp
            *  161882   162965: contig of 1084 bp in length.
FEATURES             Location/Qualifiers
     source          1. .162965
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-645I8"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .5772
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    5873. .114225
                     /note="assembly_fragment"
     misc_feature    114326. .119496
                     /note="assembly_fragment"
     misc_feature    119597. .161781
                     /note="assembly_fragment"
     misc_feature    161882. .162965
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    54235 a  33886 c  32984 g  41460 t    400 others
ORIGIN


  Query Match             29.7%;  Score 684.4;  DB 2;  Length 162965;
  Best Local Similarity   97.4%;  Pred. No. 4e-154;
  Matches  713;  Conservative    0;  Mismatches    1;  Indels   18;  Gaps    1;

Qy   1595 caaagggaacatgtctgggaaccgccgcacagccaacatcaagagcctaggttattcaga 1654
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 154404 CACAGGGAACATGTCTGGGAACCGCCGCACAGCCAACATCAAGAGCCTAGGTTATTCAGA 154463

Qy   1655 cctattctgcctgagcaaggaggacctgcgggaggtgctgagcgagtatccacaagcaca 1714
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 154464 CCTATTCTGCCTGAGCAAGGAGGACCTGCGGGAGGTGCTGAGCGAGTATCCACAAGCACA 154523

Qy   1715 gaccatcatggaggagaaaggacgtgagatcctgctgaaaatgaacaagttggacgtgaa 1774
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 154524 GACCATCATGGAGGAGAAAGGACGTGAGATCCTGCTGAAAATGAACAAGTTGGACGTGAA 154583

Qy   1775 tgctgaggcagctgagatcgccctgcaggaggccacagagtcccggctacgaggcctaga 1834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 154584 TGCTGAGGCAGCTGAGATCGCCCTGCAGGAGGCCACAGAGTCCCGGCTACGAGGCCTAGA 154643

Qy   1835 ccagcagctggatgatctacagaccaagtttgctcgcctcctggctgagctggagtccag 1894
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 154644 CCAGCAGCTGGATGATCTACAGACCAAGTTTGCTCGCCTCCTGGCTGAGCTGGAGTCCAG 154703

Qy   1895 cgcacttaagattgcttaccgcattgaacggctggagtggcagactcgagagtggccaat 1954
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 154704 CGCACTTAAGATTGCTTACCGCATTGAACGGCTGGAGTGGCAGACTCGAGAGTGGCCAAT 154763

Qy   1955 gcccgaggacctggctgaggctgatgacgagggtgagcctgaggagggaacttccaaaga 2014
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 154764 GCCCGAGGACCTGGCTGAGGCTGATGACGAGGGTGAGCCTGAGGAGGGAACTTCCAAAGA 154823

Qy   2015 tgaagagggcagggccagccaggagggaccccccaggtccagagtgaccccatccccatcc 2074
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 154824 TGAAGAGGGCAGGGCCAGCCAGGAGGGACCCCCAGGTCCAGAGTGACCCCATCCCCATCC 154883

Qy   2075 ccaggattcccacctcctagtgaatccagagttgtagtaaagcctaactgctgcaactct 2134
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 154884 CCAGGATTCCCACCTCCTAGTGAATCCAGAGTTGTAGTAAAGCCTAACTGCTGCAACTCT 154943

Qy   2135 gtcatcctgtctgcgagatcacagacacaggagcgaattggtctgtagatgcccagctag 2194
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 154944 GTCATCCTGTCTGCGAGATCACAGACACAGGAGCGAATTGGTCTGTAGATGCCCAGCTAG 155003

Qy   2195 agatataggagtttaacgcacattcagcccccacttaccagt---------- 2236
          ||||||||||||||||||||||||||||||||||||||||||
Db 155004 AGATATAGGAGTTTAACGCACATTCAGCCCCCACTTACCAGTACACACACACACACACAC 155063

Qy   2237 acacacacacacacacacacacacacatttgctcatagacctgttggccccaagactgtg 2296
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 155064 ACACACACACACACACACACACACACACATTTGCTCATAGACCTGTTGGCCCCAAGACTGTG 155123

Qy   2297 cattccatctaa 2308
          ||||||||||||
Db 155124 CATTCCATCTAA 155135


RESULT   8
AC022762
LOCUS       AC022762              186609 bp    DNA     linear   HTG 09-FEB-2002
DEFINITION  Homo sapiens chromosome 11 clone RP11-290F24 map 11, *** SEQUENCING
            IN PROGRESS ***, 1 ordered pieces.
ACCESSION   AC022762
VERSION     AC022762.6  GI:18642872
KEYWORDS    HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 186609)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-290F24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 186609)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 9, 2002 this sequence version replaced gi:18483561.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
```

```
            -------------- Project Information
                Center project name: L5842
              Center clone name: 290_F_24
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   186609: contig of 186609 bp in length.
FEATURES             Location/Qualifiers
     source          1. .186609
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-290F24"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    60812 a  39184 c  37829 g  48784 t
ORIGIN


  Query Match            29.7%;  Score 684.4;  DB 2;  Length 186609;
  Best Local Similarity  97.4%;  Pred. No. 4e-154;
  Matches  713;  Conservative    0;  Mismatches    1;  Indels   18;  Gaps    1;

Qy    1595 caaagggaacatgtctgggaaccgccgcacagccaacatcaagagcctaggttattcaga 1654
           || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  160183 CACAGGGAACATGTCTGGGAACCGCCGCACAGCCAACATCAAGAGCCTAGGTTATTCAGA 160242

Qy    1655 cctattctgcctgagcaaggaggacctgcgggaggtgctgagcgagtatccacaagcaca 1714
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  160243 CCTATTCTGCCTGAGCAAGGAGGACCTGCGGGAGGTGCTGAGCGAGTATCCACAAGCACA 160302

Qy    1715 gaccatcatggaggagaaaggacgtgagatcctgctgaaaatgaacaagttggacgtgaa 1774
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  160303 GACCATCATGGAGGAGAAAGGACGTGAGATCCTGCTGAAAATGAACAAGTTGGACGTGAA 160362

Qy    1775 tgctgaggcagctgagatcgccctgcaggaggccacagagtcccggctacgaggcctaga 1834
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  160363 TGCTGAGGCAGCTGAGATCGCCCTGCAGGAGGCCACAGAGTCCCGGCTACGAGGCCTAGA 160422

Qy    1835 ccagcagctggatgatctacagaccaagtttgctcgcctcctggctgagctggagtccag 1894
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  160423 CCAGCAGCTGGATGATCTACAGACCAAGTTTGCTCGCCTCCTGGCTGAGCTGGAGTCCAG 160482

Qy    1895 cgcacttaagattgcttaccgcattgaacggctggagtggcagactcgagagtggccaat 1954
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  160483 CGCACTTAAGATTGCTTACCGCATTGAACGGCTGGAGTGGCAGACTCGAGAGTGGCCAAT 160542

Qy    1955 gcccgaggacctggctgaggctgatgacgagggtgagcctgaggagggaacttccaaaga 2014
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  160543 GCCCGAGGACCTGGCTGAGGCTGATGACGAGGGTGAGCCTGAGGAGGGAACTTCCAAAGA 160602

Qy    2015 tgaagagggcagggccagccaggagggacccccaggtccagagtgaccccatccccatcc 2074
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  160603 TGAAGAGGGCAGGGCCAGCCAGGAGGGACCCCCAGGTCCAGAGTGACCCCATCCCCATCC 160662

Qy    2075 ccaggattcccacctcctagtgaatccagagttgtagtaaagcctaactgctgcaactct 2134
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  160663 CCAGGATTCCCACCTCCTAGTGAATCCAGAGTTGTAGTAAAGCCTAACTGCTGCAACTCT 160722

Qy    2135 gtcatcctgtctgcgagatcacagacacaggagcgaattggtctgtagatgcccagctag 2194
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  160723 GTCATCCTGTCTGCGAGATCACAGACACAGGAGCGAATTGGTCTGTAGATGCCCAGCTAG 160782

Qy    2195 agatataggagtttaacgcacattcagcccccacttaccagt------------------ 2236
           ||||||||||||||||||||||||||||||||||||||||
Db  160783 AGATATAGGAGTTTAACGCACATTCAGCCCCCACTTACCAGTACACACACACACACACAC 160842

Qy    2237 acacacacacacacacacacacacacatttgctcatagacctgttggccccaagactgtg 2296
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  160843 ACACACACACACACACACACACACACATTTGCTCATAGACCTGTTGGCCCCAAGACTGTG 160902

Qy    2297 cattccatctaa 2308
           ||||||||||||
Db  160903 CATTCCATCTAA 160914


RESULT   9
BTCNGCHAN
LOCUS       BTCNGCHAN               2641 bp    mRNA    linear   MAM 20-JUL-1995
DEFINITION  B.taurus mRNA for alpha subunit of cone and testis CNG channel.
ACCESSION   X89600
VERSION     X89600.1  GI:908823
KEYWORDS    alpha-subunit; CNG channel.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 2641)
  AUTHORS   Weyand,I., Godde,M., Frings,S., Weiner,J., Muller,F.,
            Altenhofen,W., Hatt,H. and Kaupp,U.B.
  TITLE     Cloning and functional expression of a cyclic-nucleotide-gated
            channel from mammalian sperm
  JOURNAL   Nature 368 (6474), 859-863 (1994)
  MEDLINE   94211295
REFERENCE   2  (bases 1 to 2641)
  AUTHORS   Weyand,I.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUL-1995) I. Weyand, Forschungszentrum Julich,
            Institut fur Biologische Informationsverarbeitung, Postfach 1913,
            52425 Julich, FRG
FEATURES             Location/Qualifiers
     source          1. .2641
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
     CDS             151. .2271
                     /codon_start=1
                     /product="alpha subunit of CNG-channel expressed in bovine
                     testis and retinal cone"
                     /protein_id="CAA61759.1"
                     /db_xref="GI:908824"
                     /db_xref="SWISS-PROT:Q29441"
                     /translation="MAKISTQYSHPTRTHPSVRTMDRDLDCIENGLSRTHLPCEETSS
                     ELQEGIAMETRGLAESRQSSFTSQGPTRLSRLIISLRAWSARHLHQEDQRPDSFLERF
                     RGAELQEVSSRESHVQFNVGSQEPPDRGRSAWPLARNNTNTCNNSEKDDKAKKEEKEK
                     KEEKKENPKKEEKKKDSVVMDPSSNMYYHWLTVIAVPVFYNWCLLVCRACFDELQSEH
                     LMLWLVLDYSADILYGMDMLVRARTGFLEQGLMVMDASRLWKHYTQTLHFKLDVLSLV
                     PTDLAYFKLGMNYPELRFNRLLKLARLFEFFDRTETRTNYPNMFRIGNLVLYILIIIH
                     WNACIYFAISKFIGFGTDSWVYPNVSNPEYGRLSRKYIYSLYWSTLTLTTIGETPPPV
                     KDEEYLFVVIDFLVGVLIFATIVGNVGSMISNMNASRAEFQAKIDSIKQYMQFRKVTK
                     DLETRVIRWFDYLWANKKTVDEKEVLKSLPDKLKAEIAINVHLDTLRKVRIFQDCEAG
                     LLVELVLKLRPAVFSPGDYICKKGDIGREMYIIKEGKLAVVAEDGITQFVVLGDGSYF
                     GEISILNIKGSKSGNRRTANIRSIGYSDLFCLSKDDLMEALTEYPEAKKALEEKGRQI
                     LMKDNLIDEELAKAGADPKDIEEKVEHLETSLDSLQTRFARLLAEYNATQMKVKQRLS
                     QLESQVKMGLPPDGDAPQTEASQP"
BASE COUNT      623 a    727 c    762 g    529 t
ORIGIN


  Query Match            26.8%;  Score 617.4;  DB 4;  Length 2641;
  Best Local Similarity  62.3%;  Pred. No. 4.4e-138;
  Matches  988;  Conservative    0;  Mismatches  596;  Indels    3;  Gaps    1;

Qy     407 tgtcctggacccatctggggattactactactggtggctgaacacaatggtcttcccagt 466
           |||  ||||||| ||  |  |    ||||||  .|||||||     || | | | || ||
Db     684 TGTGATGGACCCTTCCAGCAACATGTACTACCACTGGCTGACTGTCATCGCCGTGCCTGT 743

Qy     467 catgtataacctcatcatcctcgtgtgcagagcctgcttccccgacttgcagcacggtta 526
           | | || |||       | ||||||||||| ||||| |||   ||  | |||  ||  |
```

```
Db    744 CTTCTACAACTGGTGTCTGCTCGTGTGCAGGGCCTGTTTCGATGAGCTCCAGTCCGAGCA 803

Qy    527 tctggtggcctggttggtgctggactacacgagtgacctgctatacctactagacatggt 586
Db    804 CCTGATGCTTTGGCTGGTCCTGGACTACTCAGCAGACATCCTCTATGGCATGGATATGCT 863

Qy    587 ggtgcgcttccacacaggattcttggaacagggcatcctggtggtggacaagggtaggat 646
Db    864 GGTCCGAGCCCGGACAGGCTTCCTGGAGCAAGGCCTGATGGTCATGGACGCCAGCCGGCT 923

Qy    647 ctcgagtcgctacgttcgcacctggagtttcttcttggacctggcttccctgatgcccac 706
Db    924 GTGGAAGCACTACACACAGACCTTGCACTTCAAGCTGGACGTGTTGTCCCTGGTGCCCAC 983

Qy    707 agatgtggtctacgtgcggctgggcccgcacacacccaccctgaggctgaaccgctttct 766
Db    984 AGACCTGGCTTATTTTAAGCTGGGCATGAACTACCCAGAACTGAGGTTCAACCGCCTCCT 1043

Qy    767 ccgcgcgccccgcctcttcgaggccttcgaccgcacagagacccgcacagcttacccaaa 826
Db   1044 GAAGTTGGCCCGGCTCTTCGAATTCTTTGACCGCACGGAGACAAGGACCAACTACCCCAA 1103

Qy    827 tgcctttcgcattgccaagctgatgctttacatttttgtcgtcatccattggaacagctg 886
Db   1104 CATGTTCAGGATCGGGAACTTGGTCTTGTACATCCTCATCATCATCCACTGGAATGCCTG 1163

Qy    887 cctatactttgccctatcccggtacctgggcttcgggcgtgacgcatgggtgtacccgga 946
Db   1164 CATCTACTTTGCCATTTCCAAGTTCATTGGTTTCGGGACCGACTCCTGGGTCTACCCAAA 1223

Qy    947 ccccgcgcagcctggctttgagcgcctgcggcgccagtacctctatagcttttacttctc 1006
Db   1224 CGTCTCCAACCCAGAGTATGGGCGCCTCTCCAGAAAGTACATTTACAGTCTCTACTGGTC 1283

Qy   1007 cacgctgatactgactacagtgggcgatacaccgccgccagccagggaagaagagtacct 1066
Db   1284 CACCTTGACCCTGACCACCATTGGGGAGACCCCGCCCCCCGTAAAAGACGAGGAATATCT 1343

Qy   1067 cttcatggtgggcgacttcctgctggccgtcatgggtttcgccaccatcatgggtagcat 1126
Db   1344 CTTTGTGGTCATCGACTTCCTGGTGGGCGTCCTGATTTTTGCCACCATCGTGGGCAACGT 1403

Qy   1127 gagctctgtcatctacaacatgaacactgcagatgcggctttctacccagatcatgcact 1186
Db   1404 GGGCTCCATGATCTCAAACATGAATGCTTCACGGGCCGAGTTCCAGGCCAAGATCGATTC 1463

Qy   1187 ggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggagcggcgagttattga 1246
Db   1464 CATCAAGCAGTACATGCAGTTCCGCAAGGTGACCAAGGACTTGGAGACACGGGTGATCCG 1523

Qy   1247 ctggtatcagcacctgcagatcaacaagaagatgaccaacgaggtagccatcttacagca 1306
Db   1524 CTGGTTCGACTACCTGTGGGCCAATAAGAAGACAGTGGATGAGAAGGAGGTGCTCAAGAG 1583

Qy   1307 cttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctgtccactctgagccg 1366
Db   1584 CCTCCCCGACAAGCTGAAGGCCGAGATCGCCATCAACGTGCACCTGGACACCCTGAGGAA 1643

Qy   1367 ggtgcagatctttcagaactgtgaggccagcctgctggaggagctggtgctgaagctgca 1426
Db   1644 GGTCCGAATCTTCCAGGACTGCGAGGCGGGGCTGCTGGTGGAGCTGGTGCTGAAGCTGCG 1703

Qy   1427 gccccagacctactcaccaggtgaatatgtatgccgcaaaggagacattggccaagagat 1486
Db   1704 GCCGGCAGTGTTCAGCCCCGGGGACTACATCTGCAAGAAGGGGGACATCGGGAGGGAGAT 1763

Qy   1487 gtacatcatccgagagggtcaactggccgtggtggcagatgatggtatcacacagtatgc 1546
Db   1764 GTACATTATCAAGGAGGGCAAGCTGGCCGTGGTGGCCGAGGACGGCATCACCCAGTTCGT 1823

Qy   1547 tgtgctcggtgcagggctctactttggggagatcagcatcatcaacatcaaagggaacat 1606
Db   1824 GGTCCTCGGCGACGGGAGTTATTTCGGGGAGATCAGCATCTTGAACATCAAGGGGAGCAA 1883

Qy   1607 gtctgggaaccgccgcacagccaacatcaagagcctaggttattcagacctattctgcct 1666
Db   1884 GTCCGGGAACCGCCGCACGGCCAACATCAGGAGCATCGGCTACTCGGACCTGTTCTGCCT 1943

Qy   1667 gagcaaggaggacctgcgggaggtgctgagcgagtatccacaagcacagaccatcatgga 1726
Db   1944 CTCCAAGGACGACCTGATGGAGGCGCTCACCGAGTACCCCGAGGCCAAGAAGGCGCTGGA 2003

Qy   1727 ggagaaaggacgtgagatcctgctgaaaatgaacaagttggacgtgaatgctgaggcagc 1786
Db   2004 GGAGAAAGGGCGGCAGATCCTCATGAAGGACAACCTGATCGACGAGGAGCTGGCCAAGGC 2063

Qy   1787 tgagatcgccctgcaggaggccacagagtcccggctacgaggcctagaccagcagctgga 1846
Db   2064 CGGGGCAGACCCCAAGGA---CATCGAGGAGAAGGTGGAGCACCTCGAGACCTCCCTGGA 2120

Qy   1847 tgatctacagaccaagtttgctcgcctcctggctgagctggagtccagcgcacttaagat 1906
Db   2121 CTCCCTGCAGACCAGGTTTGCGCGGCTCCTGGCTGAGTACAACGCCACCCAGATGAAGGT 2180

Qy   1907 tgcttaccgcattgaacggctggagtggcagactcgagagtggccaatgcccgaggacct 1966
Db   2181 GAAGCAGCGGCTCAGCCAGTTGGAAAGCCAGGTGAAGATGGGCCTCCCGCCTGATGGCGA 2240

Qy   1967 ggctgaggctgatgacgagggtgagcc 1993
Db   2241 TGCTCCGCAAACTGAGGCCAGTCAGCC 2267
```

```
RESULT  10
BTCNG3
LOCUS       BTCNG3                  3998 bp    mRNA    linear   MAM 19-MAY-1994
DEFINITION  B.taurus CNG3 mRNA for cyclic nucleotide-gated channel.
ACCESSION   X76485
VERSION     X76485.1  GI:488728
KEYWORDS    cyclic nucleotide-gated channel.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 3998)
  AUTHORS   Biel,M., Zong,X., Distler,M., Bosse,E., Klugbauer,N., Murakami,M.,
            Flockerzi,V. and Hofmann,F.
  TITLE     Another member of the cyclic nucleotide-gated channel family,
            expressed in testis, kidney, and heart
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 91 (9), 3505-3509 (1994)
  MEDLINE   94224768
REFERENCE   2  (bases 1 to 3998)
  AUTHORS   Hofmann,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-NOV-1993) F. Hofmann, Inst. f. Pharmakologie Und
            Toxikologie, Technische Univ. Muenchen, Biedersteiner Str. 29,
            80802 Muenchen, FRG
FEATURES             Location/Qualifiers
     source          1. .3998
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /tissue_type="kidney"
                     /clone_lib="plasmid vector pSport (BRL)"
                     /dev_stage="adult"
     gene            273. .2393
                     /gene="CNG3"
     CDS             273. .2393
                     /gene="CNG3"
                     /codon_start=1
                     /product="cyclic nucleotide-gated channel 3"
                     /protein_id="CAA54023.1"
                     /db_xref="GI:488729"
                     /db_xref="SWISS-PROT:Q29441"
                     /translation="MAKISTQYSHPTRTHPSVRTMDRDLDCIENGLSRTHLPCEETSS
                     ELQEGIAMETRGLAESRQSSFTSQGPTRLSRLIISLRAWSARHLHQEDQRPDSFLERF
```

```
                     RGAELQEVSSRESHVQFNVGSQEPPDRGRSAWPLARNNTNTCNNSEKDDKAKKEEKEK
                     KEEKKENPKKEEKKKDSVVMDPSSNMYYHWLTVIAVPVFYNWCLLVCRACFDELQSEH
                     LMLWLVLDYSADILYGMDMLVRARTGFLEQGLMVMDASRLWKHYTQTLHFKLDVLSLV
                     PTDLAYFKLGMNYPELRFNRLLKLARLFEFFDRTETRTNYPNMFRIGNLVLYILIIIH
                     WNACIYFAISKFIGFGTDSWVYPNVSNPEYGRLSRKYIYSLYWSTLTLTTIGETPPPV
                     KDEEYLFVVIDFLVGVLIFATIVGNVGSMISNMNASRAEFQAKIDSIKQYMQFRKVTK
                     DLETRVIRWFDYLWANKKTVDEKEVLKSLPDKLKAEIAINVHLDTLRKVRIFQDCEAG
                     LLVELVLKLRPAVFSPGDYICKKGDIGREMYIIKEGKLAVVAEDGITQFVVLGDGSYF
                     GEISILNIKGSKSGNRRTANIRSIGYSDLFCLSKDDLMEALTEYPEAKKALEEKGRQI
                     LMKDNLIDEELAKAGADPKDIEEKVEHLETSLDSLQTRFARLLAEYNATQMKVKQRLS
                     QLESQVKMGLPPDGDAPQTEASQP"
BASE COUNT      997 a   1054 c   1050 g    897 t
ORIGIN


  Query Match           26.8%;  Score 617.4;  DB 4;  Length 3998;
  Best Local Similarity  62.3%;  Pred. No. 4.6e-138;
  Matches  988;  Conservative    0;  Mismatches  596;  Indels    3;  Gaps    1;

Qy    407 tgtcctggacccatctggggattactactactggtggctgaacacaatggtcttcccagt 466
Db    806 TGTGATGGACCCTTCCAGCAACATGTACTACCACTGGCTGACTGTCATCGCCGTGCCTGT 865

Qy    467 catgtataacctcatcatcctcgtgtgcagagcctgcttccccgacttgcagcacggtta 526
Db    866 CTTCTACAACTGGTGTCTGCTCGTGTGCAGGGCCTGTTTCGATGAGCTCCAGTCCGAGCA 925

Qy    527 tctggtggcctggttggtgctggactacacgagtgacctgctatacctactagacatggt 586
Db    926 CCTGATGCTTTGGCTGGTCCTGGACTACTCAGCAGACATCCTCTATGGCATGGATATGCT 985

Qy    587 ggtgcgcttccacacaggattcttggaacagggcatcctggtggtggacaagggtaggat 646
Db    986 GGTCCGAGCCCGGACAGGCTTCCTGGAGCAAGGCCTGATGGTCATGGACGCCAGCCGGCT 1045

Qy    647 ctcgagtcgctacgttcgcacctggagtttcttcttggacctggcttccctgatgcccac 706
Db   1046 GTGGAAGCACTACACACAGACCTTGCACTTCAAGCTGGACGTGTTGTCCCTGGTGCCCAC 1105

Qy    707 agatgtggtctacgtgcggctgggcccgcacacacccaccctgaggctgaaccgctttct 766
Db   1106 AGACCTGGCTTATTTTAAGCTGGGCATGAACTACCCAGAACTGAGGTTCAACCGCCTCCT 1165

Qy    767 ccgcgcgccccgcctcttcgaggccttcgaccgcacagagacccgcacagcttacccaaa 826
Db   1166 GAAGTTGGCCCGGCTCTTCGAATTCTTTGACCGCACGGAGACAAGGACCAACTACCCCAA 1225

Qy    827 tgcctttcgcattgccaagctgatgctttacatttttgtcgtcatccattggaacagctg 886
Db   1226 CATGTTCAGGATCGGGAACTTGGTCTTGTACATCCTCATCATCATCCACTGGAATGCCTG 1285

Qy    887 cctatactttgccctatcccggtacctgggcttcgggcgtgacgcatgggtgtacccgga 946
Db   1286 CATCTACTTTGCCATTTCCAAGTTCATTGGTTTCGGGACCGACTCCTGGGTCTACCCAAA 1345

Qy    947 ccccgcgcagcctggctttgagcgcctgcggcgccagtacctctatagcttttacttctc 1006
Db   1346 CGTCTCCAACCCAGAGTATGGGCGCCTCTCCAGAAAGTACATTTACAGTCTCTACTGGTC 1405

Qy   1007 cacgctgatactgactacagtgggcgatacaccgccgccagccagggaagaagagtacct 1066
Db   1406 CACCTTGACCCTGACCACCATTGGGGAGACCCCGCCCCCCGTAAAAGACGAGGAATATCT 1465

Qy   1067 cttcatggtgggcgacttcctgctggccgtcatgggtttcgccaccatcatgggtagcat 1126
Db   1466 CTTTGTGGTCATCGACTTCCTGGTGGGCGTCCTGATTTTTGCCACCATCGTGGGCAACGT 1525

Qy   1127 gagctctgtcatctacaacatgaacactgcagatgcggctttctacccagatcatgcact 1186
Db   1526 GGGCTCCATGATCTCAAACATGAATGCTTCACGGGCCGAGTTCCAGGCCAAGATCGATTC 1585

Qy   1187 ggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggagcggcgagttattga 1246

Db   1586 CATCAAGCAGTACATGCAGTTCCGCAAGGTGACCAAGGACTTGGAGACACGGGTGATCCG 1645

Qy   1247 ctggtatcagcacctgcagatcaacaagaagatgaccaacgaggtagccatcttacagca 1306
Db   1646 CTGGTTCGACTACCTGTGGGCCAATAAGAAGACAGTGGATGAGAAGGAGGTGCTCAAGAG 1705

Qy   1307 cttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctgtccactctgagccg 1366
Db   1706 CCTCCCCGACAAGCTGAAGGCCGAGATCGCCATCAACGTGCACCTGGACACCCTGAGGAA 1765

Qy   1367 ggtgcagatctttcagaactgtgaggccagcctgctggaggagctggtgctgaagctgca 1426
Db   1766 GGTCCGAATCTTCCAGGACTGCGAGGCGGGGCTGCTGGTGGAGCTGGTGCTGAAGCTGCG 1825

Qy   1427 gccccagacctactcaccaggtgaatatgtatgccgcaaaggagacattggccaagagat 1486
Db   1826 GCCGGCAGTGTTCAGCCCCGGGGACTACATCTGCAAGAAGGGGGACATCGGGAGGGAGAT 1885

Qy   1487 gtacatcatccgagagggtcaactggccgtggtggcagatgatggtatcacacagtatgc 1546
Db   1886 GTACATTATCAAGGAGGGCAAGCTGGCCGTGGTGGCCGAGGACGGCATCACCCAGTTCGT 1945

Qy   1547 tgtgctcggtgcagggctctactttggggagatcagcatcatcaacatcaaagggaacat 1606
Db   1946 GGTCCTCGGCGACGGGAGTTATTTCGGGGAGATCAGCATCTTGAACATCAAGGGGAGCAA 2005

Qy   1607 gtctgggaaccgccgcacagccaacatcaagagcctaggttattcagacctattctgcct 1666
Db   2006 GTCCGGGAACCGCCGCACGGCCAACATCAGGAGCATCGGCTACTCGGACCTGTTCTGCCT 2065

Qy   1667 gagcaaggaggacctgcgggaggtgctgagcgagtatccacaagcacagaccatcatgga 1726
Db   2066 CTCCAAGGACGACCTGATGGAGGCGCTCACCGAGTACCCCGAGGCCAAGAAGGCGCTGGA 2125

Qy   1727 ggagaaaggacgtgagatcctgctgaaaatgaacaagttggacgtgaatgctgaggcagc 1786
Db   2126 GGAGAAAGGGCGGCAGATCCTCATGAAGGACAACCTGATCGACGAGGAGCTGGCCAAGGC 2185

Qy   1787 tgagatcgccctgcaggaggccacagagtcccggctacgaggcctagaccagcagctgga 1846
Db   2186 CGGGGCAGACCCCAAGGA---CATCGAGGAGAAGGTGGAGCACCTCGAGACCTCCCTGGA 2242

Qy   1847 tgatctacagaccaagtttgctcgcctcctggctgagctggagtccagcgcacttaagat 1906
Db   2243 CTCCCTGCAGACCAGGTTTGCGCGGCTCCTGGCTGAGTACAACGCCACCCAGATGAAGGT 2302

Qy   1907 tgcttaccgcattgaacggctggagtggcagactcgagagtggccaatgcccgaggacct 1966
Db   2303 GAAGCAGCGGCTCAGCCAGTTGGAAAGCCAGGTGAAGATGGGCCTCCCGCCTGATGGCGA 2362

Qy   1967 ggctgaggctgatgacgagggtgagcc 1993
Db   2363 TGCTCCGCAAACTGAGGCCAGTCAGCC 2389


RESULT  11
AF065314
LOCUS       AF065314               3486 bp    mRNA    linear   PRI 27-MAY-1998
DEFINITION  Homo sapiens cone photoreceptor cGMP-gated channel alpha subunit
            (CNGA3) mRNA, complete cds.
ACCESSION   AF065314
VERSION     AF065314.1  GI:3153886
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3486)
  AUTHORS   Wissinger,B., Muller,F., Weyand,I., Schuffenhauer,S., Thanos,S.,
            Kaupp,U.B. and Zrenner,E.
  TITLE     Cloning, chromosomal localization and functional expression of the
            gene encoding the alpha-subunit of the cGMP-gated channel in human
            cone photoreceptors
```

```
  JOURNAL   Eur. J. Neurosci. 9 (12), 2512-2521 (1997)
  MEDLINE   98176633
REFERENCE   2  (bases 1 to 3486)
  AUTHORS   Wissinger,B., Muller,F., Weyand,I., Schuffenhauer,S., Thanos,S.,
            Kaupp,U.B. and Zrenner,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAY-1998) Molekulargenetisches Labor,
            Universitaets-Augenklinik Abt II, Verfuegungsgebaeude, Auf der
            Morgenstelle 15, Tuebingen D-72076, Germany
FEATURES             Location/Qualifiers
     source          1. .3486
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2q11"
     gene            1. .3486
                     /gene="CNGA3"
     CDS             40. .2124
                     /gene="CNGA3"
                     /codon_start=1
                     /product="cone photoreceptor cGMP-gated channel alpha
                     subunit"
                     /protein_id="AAC17440.1"
                     /db_xref="GI:3153887"
                     /translation="MAKINTQYSHPSRTHLKVKTSDRDLNRAENGLSRAHSSSEETSS
                     VLQPGIAMETRGLADSGQGSFTGQGIARLSRLIFLLRRWAARHVHHQDQGPDSFPDRF
                     RGAELKEVSSQESNAQANVGSQEPADRGRSAWPLAKCNTNTSNNTEEEKKTKKKDAIV
                     VDPSSNLYYRWLTAIALPVFYNWYLLICRACFDELQSEYLMLWLVLDYSADVLYVLDV
                     LVRARTGFLEQGLMVSDTNRLWQHYKTTTQFKLDVLSLVPTDLAYLKVGTNYPEVRFN
                     RLLKFSRLFEFFDRTETRTNYPNMFRIGNLVLYILIIIHWNACIYFAISKFIGFGTDS
                     WVYPNISIPEHGRLSRKYIYSLYWSTLTLTTIGETPPPVKDEEYLFVVVDFLVGVLIF
                     ATIVGNVGSMISNMNASRAEFQAKIDSIKQYMQFRKVTKDLETRVIRWFDYLWANKKT
                     VDEKEVLKSLPDKLKAEIAINVHLDTLKKVRIFQDCEAGLLVELVLKLRPTVFSPGDY
                     ICKKGDIGKEMYIINEGKLAVVADDGVTQFVVLSDGSYFGEISILNIKGSKSGNRRTA
                     NIRSIGYSDLFCLSKDDLMEALTEYPEAKKALEEKGRQILMKDNLIDEELARAGADPK
                     DLEEKVEQLGSSLDTLQTRFARLLAEYNATQMKMKQRLSQLESQVKGGGDKPLADGEV
                     PGDATKTEDKQQ"
BASE COUNT      912 a    866 c    865 g    843 t
ORIGIN
```

```
  Query Match           25.5%;  Score 588.2;  DB 9;  Length 3486;
  Best Local Similarity   61.7%;  Pred. No. 5e-131;
  Matches  954;  Conservative    0;  Mismatches  588;  Indels    3;  Gaps    1;

Qy     393 aggaagttgctgcctgtcctggacccatctggggattactactactggtggctgaacaca 452
           | ||||     |   ||  ||||||| ||  |  |    |||||| | ||||||| | |
Db     502 AAGAAGGATGCGATCGTGGTGGACCCGTCCAGCAACCTGTACTACCGCTGGCTGACCGCC 561

Qy     453 atggtcttcccagtcatgtataacctcatcatcctcgtgtgcagagcctgcttccccgac 512
           || | | | || ||| | ||||||       | ||  | ||||| ||||| |||   ||
Db     562 ATCGCCCTGCCTGTCTTCTATAACTGGTATCTGCTTATTTGCAGGGCCTGTTTCGATGAG 621

Qy     513 ttgcagcacggttatctggtggcctggttggtgctggactacacgagtgacctgctatac 572
           |||||  ||  || ||| ||   ||| |||| |||||||| ||   ||  | || ||
Db     622 CTGCAGTCCGAGTACCTGATGCTGTGGCTGGTCCTGGACTACTCGGCAGATGTCCTGTAT 681

Qy     573 ctactagacatggtggtgcgcttccacacaggattcttggaacagggcatcctggtggtg 632
           |  | ||  || | || ||    | ||||| ||  | || || ||| |  ||||
Db     682 GTCTTGGATGTGCTTGTACGAGCTCGGACAGGTTTTCTCGAGCAAGGCTTAATGGTCAGT 741

Qy     633 gacaagggtaggatctcgagtcgctacgttcgcacctggagtttcttcttggacctggct 692
           || |    ||| | | |   |  |||      ||| |   |||    ||||  ||
Db     742 GATACCAACAGGCTGTGGCAGCATTACAAGACGACCACGCAGTTCAAGCTGGATGTGTTG 801

Qy     693 tccctgatgcccacagatgtggtctacgtgcggctgggcccgcacacacccaccctgagg 752
           |||||| | ||||| ||  |||  ||| |   | ||||| |  ||   ||     |||||
Db     802 TCCCTGGTCCCCACCGACCTGGCTTACTTAAAGGTGGGCACAAACTACCCAGAAGTGAGG 861

Qy     753 ctgaaccgctttctccgcgcgccccgcctcttcgaggccttcgaccgcacagagacccgc 812
           |  |||||| | ||       |||| ||||| ||   ||| ||||||||||||||  |
Db     862 TTCAACCGCCTACTGAAGTTTTCCCGGCTCTTTGAATTCTTTGACCGCACAGAGACAAGG 921

Qy     813 acagcttacccaaatgcctttcgcattgccaagctgatgctttacatttttgtcgtcatc 872
           ||    ||||| |||   ||  | ||||  ||  || |  | ||||||  |  || |||||
Db     922 ACCAACTACCCCAATATGTTCAGGATTGGGAACTTGGTCTTGTACATTCTCATCATCATC 981

Qy     873 cattggaacagctgcctatactttgccctatcccggtacctgggcttcgggcgtgacgca 932
           || |||||   |||| | |||||||||| | |||  || | | || || |||   ||| |
Db     982 CACTGGAATGCCTGCATCTACTTTGCCATTTCCAAGTTCATTGGTTTTGGGACAGACTCC 1041

Qy     933 tgggtgtacccggaccccgcgcagcctggctttgagcgcctgcggcgccagtacctctat 992
           ||||| |||||  ||  | |     || |    || ||||||      |  ||||| | ||
Db    1042 TGGGTCTACCCAAACATCTCAATCCCAGAGCATGGGCGCCTCTCCAGGAAGTACATTTAC 1101

Qy     993 agcttttacttctccacgctgatactgactacagtgggcgatacaccgccgccagccagg 1052
           ||  | ||||  |||||  |||  || || ||  | || || || || || || |  |
Db    1102 AGTCTCTACTGGTCCACCTTGACCCTTACCACCATTGGTGAGACCCCACCCCCCGTGAAA 1161

Qy    1053 gaagaagagtacctcttcatggtgggcgacttcctgctggccgtcatgggtttcgccacc 1112
           || || ||||| |||||  |||| |  |||||| || |||  ||  ||  ||| ||||||
Db    1162 GATGAGGAGTATCTCTTTGTGGTCGTAGACTTCTTGGTGGGTGTTCTGATTTTTGCCACC 1221

Qy    1113 atcatgggtagcatgagctctgtcatctacaacatgaacactgcagatgcggctttctac 1172
           ||  |||| |   || ||||  | ||||  || |||||  |  ||  || | |||| |
Db    1222 ATTGTGGGCAATGTGGGCTCCATGATCTCGAATATGAATGCCTCACGGGCAGAGTTCCAG 1281

Qy    1173 ccagatcatgcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggag 1232
           |  |  ||     | ||| |||||||| || | |   | |||| |    | |||||
Db    1282 GCCAAGATTGATTCCATCAAGCAGTACATGCAGTTCCGCAAGGTCACCAAGGACTTGGAG 1341

Qy    1233 cggcgagttattgactggtatcagcacctgcagatcaacaagaagatgaccaacgaggta 1292
             ||| |||||    |||| | |  |||||  |  |||||||||||  |    | |||
Db    1342 ACGCGGGTTATCCGGTGGTTTGACTACCTGTGGGCCAACAAGAAGACGGTGGATGAGAAG 1401

Qy    1293 gccatcttacagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctg 1352
           |  |  |  ||  | | || ||   |||| ||| ||  | ||  |    ||||||||
Db    1402 GAGGTGCTCAAGAGCCTCCCAGACAAGCTGAAGGCTGAGATCGCCATCAACGTGCACCTG 1461

Qy    1353 tccactctgagccgggtgcagatctttcagaactgtgaggccagcctgctggaggagctg 1412
             ||| ||||    ||| |  ||||| ||| ||||||||||  | |||||| |||||||
Db    1462 GACACGCTGAAGAAGGTTCGCATCTTCCAGGACTGTGAGGCAGGGCTGCTGGTGGAGCTG 1521

Qy    1413 gtgctgaagctgcagccccagacctactcaccaggtgaatatgtatgccgcaaaggagac 1472
           |||||||||||||  |||       | |   || || || ||| | |||   || |||||
Db    1522 GTGCTGAAGCTGCGACCCACTGTGTTCAGCCCTGGGGATTATATCTGCAAGAAGGGAGAT 1581

Qy    1473 attggccaagagatgtacatcatccgagagggtcaactggccgtggtggcagatgatggt 1532
           |||||  | ||||||||||||||||  |||||  | |||||||||||||| ||||||||
Db    1582 ATTGGGAAGGAGATGTACATCATCAACGAGGGCAAGCTGGCCGTGGTGGCTGATGATGGG 1641

Qy    1533 atcacacagtatgctgtgctcggtgcagggctctactttggggagatcagcatcatcaac 1592
           ||||  |||| |  || ||| | |  ||   |||||| ||||||||||||||  | |||
Db    1642 GTCACCCAGTTCGTGGTCCTCAGCGATGGCAGCTACTTCGGGGAGATCAGCATTCTGAAC 1701

Qy    1593 atcaaagggaacatgtctgggaaccgccgcacagccaacatcaagagcctaggttattca 1652
           ||||| |||| || ||| |||||||||  | || |||||||||   ||| | || || |||
Db    1702 ATCAAGGGGAGCAAGTCGGGGAACCGCAGGACGGCCAACATCCGCAGCATTGGCTACTCA 1761

Qy    1653 gacctattctgcctgagcaaggaggacctgcgggaggtgctgagcgagtatccacaagca 1712
           ||||| ||||||||     ||||| || ||  ||||  || | |||||| ||  ||||
Db    1762 GACCTGTTCTGCCTCTCAAAGGACGATCTCATGGAGGCCCTCACCGAGTACCCCGAAGCC 1821

Qy    1713 cagaccatcatggaggagaaaggacgtgagatcctgctgaaaatgaacaagttggacgtg 1772
           |||    | ||||||||||||||||  |||||||| |||||    ||||  ||  ||
Db    1822 AAGAAGGCCCTGGAGGAGAAAGGACGGCAGATCCTGATGAAA---GACAACCTGATCGAT 1878

Qy    1773 aatgctgaggcagctgagatcgccctgcaggaggccacagagtcccggctacgaggccta 1832
           | |    |||    | |  |||    |  ||| |  |||        |        ||
Db    1879 GAGGAGCTGGCCAGGGCGGGCGCGGACCCCAAGGACCTTGAGGAGAAAGTGGAGCAGCTG 1938

Qy    1833 gaccagcagctggatgatctacagaccaagtttgctcgcctcctggctgagctggagtcc 1892
           |       |||||    || ||||||| |||||| ||||||||||||||    |  ||
Db    1939 GGGTCCTCCCTGGACACCCTGCAGACCAGGTTTGCACGCCTCCTGGCTGAGTACAACGCC 1998

Qy    1893 agcgcacttaagattgcttaccgcattgaacggctggagtggcag 1937
```

```
Db   1999 ACCCAGATGAAGATGAAGCAGCGTCTCAGCCAACTGGAAAGCCAG 2043


RESULT 12
MMU243933
LOCUS       MMU243933               2781 bp    mRNA    linear   ROD 25-MAY-2000
DEFINITION  Mus musculus mRNA for cone photoreceptor cyclic nucleotide-gated
            channel alpha-subunit.
ACCESSION   AJ243933
VERSION     AJ243933.1  GI:7688040
KEYWORDS    CNG3 gene; cyclic nucleotide-gated channel alpha-subunit.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2781)
  AUTHORS   Hirano,A.A., Hack,I., Wassle,H. and Duvoisin,R.M.
  TITLE     Cloning and immunocytochemical localization of a cyclic
            nucleotide-gated channel alpha-subunit to all cone photoreceptors
            in the mouse retina
  JOURNAL   J. Comp. Neurol. 421 (1), 80-94 (2000)
  MEDLINE   20273944
REFERENCE   2  (bases 1 to 2781)
  AUTHORS   Hirano,A.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-1999) Hirano A.A., Neuroanatomie,
            Max-Planck-Institut fuer Hirnforschung, Deutschordenstrasse 46,
            D-60528 Frankfurt a.M., GERMANY
FEATURES             Location/Qualifiers
     source          1. .2781
                     /organism="Mus musculus"
                     /plasmid="pNC-2"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /tissue_type="retina"
                     /dev_stage="adult"
     gene            380. .2275
                     /gene="CNG3"
     CDS             380. .2275
                     /gene="CNG3"
                     /function="phototransduction in cone photoreceptors"
                     /codon_start=1
                     /evidence=experimental
                     /product="cyclic nucleotide-gated channel alpha-subunit"
                     /protein_id="CAB89685.1"
                     /db_xref="GI:7688041"
                     /translation="MAKVNTQCSQPSPTQLSIKNADRDLDHVENGLGRVSRLIISIRA
                     WASRHLHDEDQTPDSFLDRFHGSELKEVSTRESNAQPNPGEQKPPDGGEGRKEEPIVV
                     DPSSNIYCRWLTAIALPVFYNWCLLVCRACFDELQSEHLTLWLVLDYSADVLYVLDML
                     VRARTGFLEQGLMVRDTKRLWKHYTKTLHFKLDILSLIPTDLAYLKLGVNYPELRFNR
                     LLKFSRLFEFFDRTETRTNYPNVFRIGNLVLYTLIIIHWNACIYFAISKFIGFGTDSW
                     VYPNTSKPEYARLSRKYIYSLYWSTLTLTTIGETPPPVKDEEYLFVVIDFLVGILIFA
                     TIVGNVGSMISNMNAPRVEFQAKIDSVKQYMQFRKVTKDLETRVIRWFDYLWANRKTV
                     DEKEVLKNLPDKLKAEIAINVHLDTLKKVRIFQDCEAGLLVELVLKLRPTVFSPGDYI
                     CKKGDIGREMYIIKEGKLAVVADDGVTQFVVLSDGSYFGEISILNIKGSKSGNRRTAN
                     IRSIGYSDLFCLSKDDLMEALTEYPDAKRALEEKGRQILMKDNLIDEDLVAARVDTRD
                     VEEKVEYLESSLDILQTRFARLLAEYSASQMKLKQRLTRLESQMNRRCCGFSPDRENS
                     EDASKTD"
BASE COUNT      672 a    707 c    743 g    659 t
ORIGIN


  Query Match           25.3%;  Score 584.8;  DB 10;  Length 2781;
  Best Local Similarity   61.4%;  Pred. No. 3.2e-130;
  Matches  975;  Conservative    0;  Mismatches  607;  Indels    6;  Gaps    2;

Qy     354 aagacaacagagtccagtcccccagccccatccaaggccaggaagttgctgcc---tgtc 410
Db     623 AACCCAGGAGAACAGAAGCCACCAGACGGAGGGGAAGGCAGGAAGGAGGAGCCCATTGTG 682

Qy     411 ctggacccatctggggattactactactggtggctgaacacaatggtcttcccagtcatg 470
Db     683 GTGGACCCCTCCAGCAACATCTACTGCCGCTGGCTGACTGCCATCGCGCTCCCAGTCTTC 742

Qy     471 tataacctcatcatcctcgtgtgcagagcctgcttccccgacttgcagcacggttatctg 530
Db     743 TATAACTGGTGTCTACTTGTATGCAGGGCCTGTTTTGATGAGCTACAATCAGAGCACCTG 802

Qy     531 gtggcctggttggtgctggactacacgagtgacctgctatacctactagacatggtggtg 590
Db     803 ACGCTGTGGCTGGTCTTGGACTACTCTGCAGATGTCCTTTATGTTTTAGACATGCTGGTT 862

Qy     591 cgcttccacacaggattcttggaacagggcatcctggtggtggacaagggtaggatctcg 650
Db     863 CGAGCCCGGACAGGTTTCCTTGAGCAAGGCCTAATGGTCAGAGATACCAAGAGGCTGTGG 922

Qy     651 agtcgctacgttcgcacctggagtttcttcttggacctggcttccctgatgcccacagat 710
Db     923 AAACATTACACAAAGACCTTGCACTTCAAGCTGGACATCCTGTCTCTCATCCCCACAGAC 982

Qy     711 gtggtctacgtgcggctgggcccgcacacacccaccctgaggctgaaccgctttctccgc 770
Db     983 CTGGCTTATTTAAAGTTGGGCGTGAACTACCCAGAACTGAGGTTCAACCGCCTCCTGAAG 1042

Qy     771 gcgccccgcctcttcgaggccttcgaccgcacagagacccgcacagcttacccaaatgcc 830
Db    1043 TTCTCTCGGCTCTTTGAATTCTTTGACCGCACAGAGACAAGGACCAACTACCCCAACGTG 1102

Qy     831 tttcgcattgccaagctgatgctttacatttttgtcgtcatccattggaacagctgccta 890
Db    1103 TTCAGGATAGGGAACTTGGTCCTTTACACACTCATCATCATCCACTGGAATGCCTGCATC 1162

Qy     891 tactttgccctatcccggtacctgggcttcgggcgtgacgcatgggtgtacccggacccc 950
Db    1163 TACTTTGCCATTTCCAAGTTCATTGGTTTTGGGACAGACTCCTGGGTCTATCCAAACACC 1222

Qy     951 gcgcagcctggctttgagcgcctgcggcgccagtacctctatagcttttacttctccacg 1010
Db    1223 TCCAAGCCGGAGTATGCACGACTCTCCCGGAAGTACATTTACAGTCTCTACTGGTCCACC 1282

Qy    1011 ctgatactgactacagtgggcgatacaccgccgccagccagggaagaagagtacctcttc 1070
Db    1283 CTGACCCTGACCACCATTGGTGAGACCCCACCCCCCGTGAAAGATGAAGAGTATCTCTTT 1342

Qy    1071 atggtgggcgacttcctgctggccgtcatgggtttcgccaccatcatgggtagcatgagc 1130
Db    1343 GTGGTCATCGACTTCCTGGTGGGTATCCTGATCTTTGCCACCATAGTGGGCAACGTAGGC 1402

Qy    1131 tctgtcatctacaacatgaacactgcagatgcggctttctacccagatcatgcactggtg 1190
Db    1403 TCCATGATTTCCAACATGAACGCTCCCCGGGTAGAGTTCCAGGCTAAGATCGACTCCGTC 1462

Qy    1191 aagaagtacatgaagctgcagcacgtcaaccgcaagctggagcggcgagttattgactgg 1250
Db    1463 AAGCAGTACATGCAGTTCCGGAAGGTAACCAAGGACTTGGAGACTCGGGTTATCCGGTGG 1522

Qy    1251 tatcagcacctgcagatcaacaagaagatgaccaacgaggtagccatcttacagcacttg 1310
Db    1523 TTTGACTATCTGTGGGCCAACAGGAAGACGGTGGATGAAAAGGAAGTGCTCAAAAACCTC 1582

Qy    1311 cctgagcggctgcgggcagaagtggctgtgtctgtgcacctgtccactctgagccgggtg 1370
Db    1583 CCGGACAAGCTGAAGGCTGAGATCGCCATCAACGTGCACCTGGACACGCTGAAGAAGGTC 1642

Qy    1371 cagatctttcagaactgtgaggccagcctgctggaggagctggtgctgaagctgcagccc 1430
Db    1643 CGAATCTTCCAGGACTGCGAGGCGGGGCTGCTGGTGGAGCTGGTGCTGAAGCTCCGCCCC 1702

Qy    1431 cagacctactcaccaggtgaatatgtatgccgcaaaggagacattggccaagagatgtac 1490
Db    1703 ACTGTGTTCAGCCCCGGGGACTACATATGCAAAAAGGGGGACATTGGAAGAGAGATGTAC 1762

Qy    1491 atcatccgagagggtcaactggccgtggtggcagatgatggtatcacacagtatgctgtg 1550
Db    1763 ATCATCAAGGAGGGCAAGCTGGCTGTGGTGGCTGATGACGGGGTCACCCAGTTTGTGGTC 1822
```

```
Qy   1551 ctcggtgcagggctctactttggggagatcagcatcatcaacatcaaagggaacatgtct 1610
          ||| |||  ||    ||||||||||||||| |||||| | ||||| || |||| || ||||
Db   1823 CTCAGTGACGGCAGTTACTTTGGGGAGATTAGCATCTTAAACATTAAGGGGAGCAAGTCT 1882

Qy   1611 gggaaccgccgcacagccaacatcaagagcctaggttattcagacctattctgcctgagc 1670
          |||||||||| | || |||||||||| ||| | || || |||||||| ||||||||   |
Db   1883 GGGAACCGCAGGACGGCCAACATCAGGAGTATCGGCTACTCAGACCTGTTCTGCCTCTCC 1942

Qy   1671 aaggaggacctgcgggaggtgctgagcgagtatccacaagcacagaccatcatggaggag 1730
          ||||| |||||   ||| |  || | |||||| ||| | ||  |||      |||||||
Db   1943 AAGGATGACCTAATGGAAGCCCTCACCGAGTACCCAGACGCTAAGAGGGCTCTGGAGGAA 2002

Qy   1731 aaaggacgtgagatcctgctgaaaatgaacaagttggacgtgaatgctgaggcagctgag 1790
          || || ||| |||| ||| ||||    |||    | || | | |    |  || ||   |
Db   2003 AAGGGCCGTCAGATTCTGATGAAGGACAACCTAATTGATGAGGACCTAGTCGCGGCCAGG 2062

Qy   1791 atcgccctgcaggaggccacagagtcccggctacgaggcctagaccagcagctggatgat 1850
          | |       |||   |  |||     | |      ||| ||   |   |||||
Db   2063 GTAGATACCAGGGA---CGTTGAGGAGAAGGTGGAGTACCTGGAGTCGTCCCTGGACATC 2119

Qy   1851 ctacagaccaagtttgctcgcctcctggctgagctggagtccagcgcacttaagattgct 1910
          || ||||| | |||||| || ||||||||||||     ||  |    | ||| |
Db   2120 CTGCAGACGAGGTTTGCCCGACTCCTGGCTGAGTACAGTGCCTCCCAGATGAAGCTGAAA 2179

Qy   1911 taccgcattgaacggctggagtggcaga 1938
          | ||  |    ||||||||| | ||||
Db   2180 CAGCGGCTCACTCGGCTGGAGAGCCAGA 2207


RESULT  13
BTCAMPGC
LOCUS       BTCAMPGC                3166 bp    mRNA    linear   MAM 11-MAR-1993
DEFINITION  B.taurus mRNA for cAMP-gated channel.
ACCESSION   X55010
VERSION     X55010.1  GI:287743
KEYWORDS    cAMP-grated channel.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 3166)
  AUTHORS   Ludwig,J., Margalit,T., Eismann,E., Lancet,D. and Kaupp,U.B.
  TITLE     Primary structure of cAMP-gated channel from bovine olfactory
            epithelium
  JOURNAL   FEBS Lett. 270 (1-2), 24-29 (1990)
  MEDLINE   91032022
FEATURES             Location/Qualifiers
     source          1. .3166
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /cell_type="olfactory epithelium"
     CDS             418. .2409
                     /codon_start=1
                     /product="cAMP-gated channel"
                     /protein_id="CAA38754.1"
                     /db_xref="GI:287744"
                     /db_xref="SWISS-PROT:Q03041"
                     /translation="MTEKANGVKSSPANNHNHHAPPAIKASGKDDHRASSRPQSAAAD
                     DTSSELQQLAEMDAPQQRRGGFRRIARLVGVLREWAYRNFREEEPRPDSFLERFRGPE
                     LHTVTTQQGDGKGDKDGEGKGTKKKFELFVLDPAGDWYYRWLFLIALPVLYNWCLLVA
                     RACFSDLQKGYYIVWLVLDYVSDVVYIADLFIRLRTGFLEQGLLVKDTKKLRDNYIHT
                     MQFKLDVASIIPTDLIYFAVGIHNPEVRFNRLLHFARMFEFFDRTETRTSYPNIFRIS
                     NLILYILIIIHWNACIYYAISKSIGFGVDTWVYPNITDPEYGYLSREYIYCLYWSTLT
                     LTTIGETPPPVKDEEYLFVIFDFLIGVLIFATIVGNVGSMISNMNATRAEFQAKIDAV
                     KHYMQFRKVSKEMEAKVIRWFDYLWTNKKSVDEREVLKNLPAKLRAEIAINVHLSTLK
                     KVRIFQDCEAGLLVELVLKLRPQVFSPGDYICRKGDIGKEMYIIKEGKLAVVADDGVT
                     QYALLSAGSCFGEISILNIKGSKMGNRRTANIRSLGYSDLFCLSKDDLMEAVTEYPDA
                     KRVLEERGREILMKEGLLDENEVAASMEVDVQEKLEQLETNMDTLYTRFARLLAEYTG
                     AQQKLKQRITVLETKMKQNNEDDSLSDGMNSPEPPAEKP"
BASE COUNT      734 a    831 c    883 g    718 t
ORIGIN
```

```
  Query Match           25.3%;  Score 583.2;  DB 4;  Length 3166;
  Best Local Similarity   61.5%;  Pred. No. 7.9e-130;
  Matches  955;  Conservative    0;  Mismatches  593;  Indels    6;  Gaps    1;

Qy    385 ccaaggccaggaagttgctgcctgtcctggacccatctggggattactactactggtggc 444
          |||||   | |     ||   |||| |||||||| | ||||| |  |||||| | ||||
Db    788 CCAAGAAGAAGTTTGAACTCTTTGTCTTGGACCCAGCCGGGGACTGGTACTACCGCTGGC 847

Qy    445 tgaacacaatggtcttcccagtcatgtataacctcatcatcctcgtgtgcagagcctgct 504
          |       || | ||| || ||| | || |||     | |  | |||  ||||||||||
Db    848 TTTTTCTCATTGCCTTGCCCGTCCTCTACAACTGGTGCCTATTGGTGGCCAGAGCCTGCT 907

Qy    505 tccccgacttgcagcacggttatctggtggcctggttggtgctggactacacgagtgacc 564
          ||   ||| ||||| | || ||      | |  ||| ||||||||||| |||      ||
Db    908 TCAGTGACCTGCAGAAAGGCTACTACATAGTGTGGCTGGTGCTGGATTACGTCTCAGATG 967

Qy    565 tgctatacctactagacatggtggtgcgcttccacacaggattcttggaacagggcatcc 624
          || | ||| |   |||| |  |  | ||  | | |||||| |||||||| |||||  | |
Db    968 TGGTCTACATCGCAGACCTCTTCATCCGACTGCGCACAGGTTTCTTGGAGCAGGGGCTAC 1027

Qy    625 tggtggtggacaagggtaggatctcgagtcgctacgttcgcacctggagtttcttcttgg 684
          |||||   ||||      | | |   |     |||| | | |||| |   ||     |||
Db    1028 TGGTGAAAGACACCAAGAAGTTGCGGGACAACTACATCCACACCATGCAGTTTAAGCTGG 1087

Qy    685 acctggcttccctgatgcccacagatgtggtctacgtgcggctgggcccgcacacaccca 744
          |  |||| ||| | || || ||||| || ||||  |     ||||    || |  ||
Db    1088 ATGTGGCCTCCATCATCCCTACAGACCTGATCTATTTTGCTGTGGGGATCCATAACCCTG 1147

Qy    745 ccctgaggctgaaccgctttctccgcgcgccccgcctcttcgaggccttcgaccgcacag 804
            || |  | |||||| | || | |     ||||| | || |||  ||| ||||||||  |
Db    1148 AGGTGCGCTTCAACCGCCTGCTACACTTTGCCCGCATGTTTGAGTTCTTTGACCGCACTG 1207

Qy    805 agacccgcacagcttacccaaatgcctttcgcattgccaagctgatgctttacatttttg 864
          |||| |||||    ||||| ||   ||| || ||   ||| ||||| || |||||  ||
Db    1208 AGACACGCACCAGCTACCCCAACATCTTCCGAATAAGCAACCTGATCCTCTACATCTTGA 1267

Qy    865 tcgtcatccattggaacagctgcctatactttgccctatcccggtacctgggcttcgggc 924
          || |||| || |||||   |||| | |||| |||| | |||  || | | |||||| |||
Db    1268 TCATCATTCACTGGAATGCCTGCATCTACTATGCCATCTCCAAGTCCATCGGCTTTGGGG 1327

Qy    925 gtgacgcatgggtgtacccggaccccgcgcagcctggctttgagcgcctgcggcgccagt 984
            ||| | ||||| |||||  ||  | |  | ||||  | ||    ||||    |  |||
Db    1328 TAGACACCTGGGTTTACCCCAACATCACTGACCCTGAGTATGGCTACCTGTCTAGGGAGT 1387

Qy    985 acctctatagctttacttctccacgctgatactgactacagtgggcgatacaccgccgc 1044
          || ||||| || |||||| || || ||||  || || ||  | || || ||||| || |
Db    1388 ACATCTATTGCCTTTACTGGTCTACACTGACCCTCACCACCATTGGGGAGACACCACCCC 1447

Qy    1045 cagccagggaagaagagtacctcttcatggtgggcgacttcctgctggccgtcatgggtt 1104
           | |  | ||| || |||||||| ||  |  |    |||||||||  | |  ||| |    |
Db    1448 CTGTAAAGGATGAGGAGTACCTGTTTGTCATCTTTGACTTCCTGATTGGTGTCCTCATCT 1507

Qy    1105 tcgccaccatcatgggtagcatgagctctgtcatctacaacatgaacactgcagatgcgg 1164
           | |||||||| |||| |   || ||||  | |||| |||||||||  |  |    || |
Db    1508 TTGCCACCATCGTGGGAAATGTGGGCTCCATGATCTCCAACATGAATGCCACCCGGGCTG 1567

Qy    1165 ctttctacccagatcatgcactggtgaagaagtacatgaagctgcagcacgtcaaccgca 1224
           |||  |  |  |   ||     || ||  | || ||| || | |   | |||| |
Db    1568 AGTTCCAGGCCAAGATTGATGCTGTCAAACATTATATGCAGTTCCGAAAGGTCAGCAAGG 1627

Qy    1225 agctggagcggcgagttattgactggtatcagcacctgcagatcaacaagaagatgacca 1284
           || ||||       || |||   |||| | |  || ||  || ||| |||||||
Db    1628 AGATGGAAGCCAAGGTCATTAGGTGGTTTGACTACTTGTGGACCAATAAGAAGAGTGTAG 1687

Qy    1285 acgaggtagccatcttacagcacttgcctgagcggctgcgggcagaagtggctgtgtctg 1344
           | |||  ||   || |  |  || |||| |   |||  |||| ||  | ||  |   |
Db    1688 ATGAGCGAGAAGTCCTCAAAAACCTGCCAGCAAAGCTCAGGGCTGAGATAGCCATCAACG 1747

Qy    1345 tgcacctgtccactctgagccgggtgcagatctttcagaactgtgaggccagcctgctgg 1404
           | |||||||||| || |      |||| ||||||||| ||||||||||  |||||||||
Db    1748 TCCACCTGTCCACACTCAAGAAAGTGCGCATCTTTCAGGACTGTGAGGCTGGCCTGCTGG 1807
```

```
Qy   1405 aggagctggtgctgaagctgcagccccagacctactcaccaggtgaatatgtatgccgca 1464
Db   1808 TGGAACTGGTATTAAAGCTCCGGCCTCAGGTCTTTAGCCCTGGGGACTACATTTGCCGCA 1867

Qy   1465 aaggagacattggccaagagatgtacatcatccgagagggtcaactggccgtggtggcag 1524
Db   1868 AGGGGGATATTGGGAAGGAGATGTACATAATCAAGGAGGGAAAATTGGCAGTGGTGGCTG 1927

Qy   1525 atgatggtatcacacagtatgctgtgctcggtgcagggctctactttggggagatcagca 1584
Db   1928 ATGACGGTGTCACTCAGTATGCCCTGCTCTCGGCTGGGAGTTGCTTTGGAGAGATCAGTA 1987

Qy   1585 tcatcaacatcaaagggaacatgtctgggaaccgccgcacagccaacatcaagagcctag 1644
Db   1988 TCCTTAATATTAAGGGCAGCAAAATGGGCAATCGGCGCACAGCCAACATCCGCAGTCTTG 2047

Qy   1645 gttattcagacctattctgcctgagcaaggaggacctgcgggaggtgctgagcgagtatc 1704
Db   2048 GCTACTCTGATCTGTTCTGCTTGTCCAAGGATGATCTTATGGAAGCTGTGACTGAGTACC 2107

Qy   1705 cacaagcacagaccatcatggaggagaaaggacgtgagatcctgctgaaaatgaacaagt 1764
Db   2108 CTGATGCCAAGAGGGTCTTGGAGGAGAGAGGCCGGGAGATTCTGATGAAGGAGGGCTTGT 2167

Qy   1765 tggacgtgaatgctgaggcagctgagatcgccctgcaggaggccacagagtcccggctac 1824
Db   2168 TGGATGAGAATGAGGTGGCAGC------CAGCATGGAGGTAGATGTGCAGGAAAAGCTAG 2221

Qy   1825 gaggcctagaccagcagctggatgatctacagaccaagtttgctcgcctcctggctgagc 1884
Db   2222 AACAGCTGGAGACCAACATGGACACCTTGTACACTCGTTTTGCCCGCCTGCTGGCCGAGT 2281

Qy   1885 tggagtccagcgcacttaagattgcttaccgcattgaacggctggagtggcaga 1938
Db   2282 ACACGGGAGCCCAGCAGAAGCTCAAGCAGCGCATCACAGTTTTGGAAACGAAGA 2335


RESULT 14
AX179746
LOCUS       AX179746                1995 bp    DNA     linear   PAT 06-AUG-2001
DEFINITION  Sequence 34 from Patent WO0146258.
ACCESSION   AX179746
VERSION     AX179746.1  GI:15132110
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1995)
  AUTHORS   Baughn,M.R., Burford,N., Au-Young,J., Lu,D.A., Yang,J., Reddy,R.,
            Lal,P., Hillman,J.L., Azimzai,Y., Yue,H., Nguyen,D.B., Yao,M.G.,
            Gandhi,A.R., Tang,Y.T. and Khan,F.A.
  TITLE     Transporters and ion channels
  JOURNAL   Patent: WO 0146258-A 34 28-JUN-2001;
            Incyte Genomics, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .1995
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /note="Incyte ID No: 7472002CB1"
BASE COUNT      511 a    511 c    532 g    441 t
ORIGIN


  Query Match           25.1%;  Score 578.6;  DB 6;  Length 1995;
  Best Local Similarity   60.7%;  Pred. No. 9.9e-129;
  Matches  966;  Conservative    0;  Mismatches  619;  Indels    6;  Gaps    1;

Qy    402 ctgcctgtcctggacccatctggggattactactactggtggctgaacacaatggtcttc 461
Db    388 CTATTTGTCTTGGACCCAGCTGGGGATTGGTACTACTGCTGGCTATTTGTCATTGCCATG 447

Qy    462 ccagtcatgtataacctcatcatcctcgtgtgcagagcctgcttccccgacttgcagcac 521
Db    448 CCCGTCCTTTACAACTGGTGCCTGCTGGTGGCCAGAGCCTGCTTCAGTGACCTACAGAAA 507

Qy    522 ggttatctggtggcctggttggtgctggactacacgagtgacctgctatacctactagac 581
Db    508 GGCTACTACCTGGTGTGGCTGGTGCTGGATTATGTCTCAGATGTGGTCTACATTGCGGAC 567

Qy    582 atggtggtgcgcttccacacaggattcttggaacagggcatcctggtggtggacaagggt 641
Db    568 CTCTTCATCCGATTGCGCACAGGTTTCCTGGAGCAGGGGCTGCTGGTCAAAGATACCAAG 627

Qy    642 aggatctcgagtcgctacgttcgcacctggagtttcttcttggacctggcttccctgatg 701
Db    628 AAACTGCGAGACAACTACATCCACACCCTGCAGTTCAAGCTGGATGTGGCTTCCATCATC 687

Qy    702 cccacagatgtggtctacgtgcggctgggcccgcacacacccaccctgaggctgaaccgc 761
Db    688 CCCACTGACCTGATCTATTTTGCTGTGGACATCCACAGCCCTGAGGTGCGCTTCAACCGC 747

Qy    762 tttctccgcgcgccccgcctcttcgaggccttcgaccgcacagagacccgcacagcttac 821
Db    748 CTGCTGCACTTTGCCCGCATGTTTGAGTTCTTTGACCGGACAGAGACACGCACCAACTAC 807

Qy    822 ccaaatgcctttcgcattgccaagctgatgctttacatttttgtcgtcatccattggaac 881
Db    808 CCTAACATCTTCCGCATCAGCAACCTTGTCCTCTACATCTTGGTCATCATCCACTGGAAT 867

Qy    882 agctgcctatactttgccctatcccggtacctgggcttcgggcgtgacgcatgggtgtac 941
Db    868 GCCTGCATCTATTATGCCATCTCCAAATCCATAGGCTTTGGGGTCGACACCTGGGTTTAC 927

Qy    942 ccggaccccgcgcagcctggctttgagcgcctgcggcgccagtacctctatagcttttac 1001
Db    928 CCAAACATCACTGACCCTGAGTATGGCTACCTGGCTAGGGAATACATCTATTGCCTTTAC 987

Qy   1002 ttctccacgctgatactgactacagtgggcgatacaccgccgccagccagggaagaagag 1061
Db    988 TGGTCCACACTGACTCTCACTACCATTGGGGAGACACCACCCCCTGTAAAGGATGAGGAG 1047

Qy   1062 tacctcttcatggtgggcgacttcctgctggccgtcatgggtttcgccaccatcatgggt 1121
Db   1048 TACCTATTTGTCATCTTTGACTTCCTGATTGGCGTCCTCATCTTTGCCACCATCGTGGGA 1107

Qy   1122 agcatgagctctgtcatctacaacatgaacactgcagatgcggctttctacccagatcat 1181
Db   1108 AATGTGGGCTCCATGATCTCCAACATGAATGCCACCCGGGCAGAGTTCCAGGCTAAGATC 1167

Qy   1182 gcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggagcggcgagtt 1241
Db   1168 GATGCCGTGAAACACTACATGCAGTTCCGAAAGGTCAGCAAGGGGATGGAAGCCAAGGTC 1227

Qy   1242 attgactggtatcagcacctgcagatcaacaagaagatgaccaacgaggtagccatctta 1301
Db   1228 ATTAGGTGGTTTGACTACTTGTGGACCAATAAGAAGACAGTGGATGAGCGAGAAATTCTC 1287

Qy   1302 cagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctgtccactctg 1361
Db   1288 AAGAATCTGCCAGCCAAGCTCAGGGCTGAGATAGCCATCAATGTCCACTTGTCCACACTC 1347

Qy   1362 agccgggtgcagatctttcagaactgtgaggccagcctgctggaggagctggtgctgaag 1421
Db   1348 AAGAAAGTGCGCATCTTCCATGATTGTGAGGCTGGCCTGCTGGTAGAGCTGGTACTGAAA 1407

Qy   1422 ctgcagccccagacctactcaccaggtgaatatgtatgccgcaaaggagacattggccaa 1481
Db   1408 CTCCGTCCTCAGGTCTTCAGTCCTGGGGATTACATTTGCCGCAAAGGGGACATCGGCAAG 1467

Qy   1482 gagatgtacatcatccgagagggtcaactggccgtggtggcagatgatggtatcacacag 1541
Db   1468 GAGATGTACATCATTAAGGAGGGCAAACTGGCAGTGGTGGCTGATGATGGTGTGACTCAG 1527

Qy   1542 tatgctgtgctcggtgcagggctctactttggggagatcagcatcatcaacatcaaaggg 1601
```

```
Db    1528 TATGCTCTGCTGTCGGCTGGAAGCTGCTTTGGCGAGATCAGTATCCTTAACATTAAGGGC 1587

Qy    1602 aacatgtctgggaaccgccgcacagccaacatcaagagcctaggttattcagacctattc 1661
Db    1588 AGTAAAATGGGCAATCGACGCACAGCTAATATCCGCAGCCTGGGCTACTCAGATCTCTTC 1647

Qy    1662 tgcctgagcaaggaggacctgcgggaggtgctgagcgagtatccacaagcacagaccatc 1721
Db    1648 TGCTTGTCCAAGGATGATCTTATGGAAGCTGTGACTGAGTACCCTGATGCCAAGAAAGTC 1707

Qy    1722 atggaggagaaaggacgtgagatcctgctgaaaatgaacaagttggacgtgaatgctgag 1781
Db    1708 CTAGAAGAGAGGGGTCGGGAGATCCTCATGAAGGAGGGACTGCTGGATGAGAACGAAGTG 1767

Qy    1782 gcagctgagatcgccctgcaggaggccacagagtcccggctacgaggcctagaccagcag 1841
Db    1768 GCA------ACCAGCATGGAGGTCGACGTGCAGGAGAAGCTAGGGCAGCTGGAGACCAAC 1821

Qy    1842 ctggatgatctacagaccaagtttgctcgcctcctggctgagctggagtccagcgcactt 1901
Db    1822 ATGGAAACCTTGTACACTCGCTTTGGCCGCCTGCTGGCTGAGTACACGGGGGCCCAGCAG 1881

Qy    1902 aagattgcttaccgcattgaacggctggagtggcagactcgagagtggccaatgcccgag 1961
Db    1882 AAGCTCAAGCAGCGCATCACAGTTCTGGAAACCAAGATGAAACAGAACAATGAAGATGAC 1941

Qy    1962 gacctggctgaggctgatgacgagggtgagc 1992
Db    1942 TACCTGTCTGATGGGATGAACAGCCCTGAGC 1972


RESULT 15
OCRACNG
LOCUS       OCRACNG                 2669 bp    mRNA    linear   MAM 08-DEC-1993
DEFINITION  O.cuniculus rACNG mRNA for aorta CNG channel.
ACCESSION   X59668
VERSION     X59668.1  GI:433959
KEYWORDS    CNG channel.
SOURCE      rabbit.
  ORGANISM  Oryctolagus cuniculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
REFERENCE   1  (bases 1 to 2669)
  AUTHORS   Biel,M., Altenhofen,W., Hullin,R., Ludwig,J., Freichel,M.,
            Flockerzi,V., Dascal,N., Kaupp,U.B. and Hofmann,F.
  TITLE     Primary structure and functional expression of a cyclic
            nucleotide-gated channel from rabbit aorta
  JOURNAL   FEBS Lett. 329 (1-2), 134-138 (1993)
  MEDLINE   93359035
REFERENCE   2  (bases 1 to 2669)
  AUTHORS   Hofmann,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-1993) F. Hofmann, Institut f. Pharmakologie &
            Toxikologie, Technische Universitaet Muenchen, Biedersteiner Str.
            29, 80802 Muenchen, FRG
FEATURES             Location/Qualifiers
     source          1. .2669
                     /organism="Oryctolagus cuniculus"
                     /db_xref="taxon:9986"
                     /tissue_type="aorta"
                     /clone_lib="plasmid vector pSport (BRL)"
                     /dev_stage="adult"
     gene            153. .2351
                     /gene="rACNG"
     CDS             153. .2351
                     /gene="rACNG"
                     /codon_start=1
                     /product="aorta CNG channel (rACNG)"
                     /protein_id="CAA42201.1"
                     /db_xref="GI:433960"
                     /db_xref="SWISS-PROT:Q28718"
                     /translation="MSSWRSCARAPLSGSAWRRSAATRRSRRCLKTKRKRWSSGKGTP
                     MQSTQCETRRRAQTPCESTGHTWRMTEKSNGVKSSPANNHNNHVPATIKANGKDESRT
                     RSRPQSAADDDTSSELQRLAEMDAPQQRRGGFRRIVRLVGVIRQWANRNFREEEARPD
                     SFLERFRGPELQTVTTQQGDGKGDKDGDGKGTKKKFELFVLDPAGDWYYRWLFVIAMP
                     VLYNWCLLVARACFSDLQRGYFLVWLVLDYFSDVVYIADLFIRLRTGFLEQGLLVKDP
                     KKLRDNYIHTLQFKLDVASIIPTDLIYFAVGIHNPELRFNRLLHFARMFEFFDRTETR
                     TSYPNIFRISNLVLYILVIIHWNACIYYAISKSIGFGVDTWVYPNITDPEYGYLAREY
                     IYCLYWSTLTLTTIGETPPPVKDEEYLFVIFDFLIGVLIFATIVGNVGSMISNMNATR
                     AEFQAKIDAVKHYMQFRKVSKEMEAKVIKWFDYLWTNKKTVDEREVLKNLPAKLRAEI
                     AINVHLSTLKKVRIFQDCEAGLLVELVLKLRPQVFSPGDYICRKGDIGKEMYIIKEGK
                     LAVVADDGVTQYALLSAGSCFGEISILNIKGSKMGNRRTANIRSLGYSDLFCLSKDDL
                     MEAVTEYPDAKKVLEERGREILMKEGLLDENEVAASMEVDVQEKLKQLETNMETLYTR
                     FGRLLAEYTGAQQKLKQRITVLEVKMKQNTEDDYLSDGMNSPEPAAAEQP"
BASE COUNT      682 a    639 c    742 g    606 t
ORIGIN


  Query Match             24.8%;  Score 572;  DB 4;  Length 2669;
  Best Local Similarity   60.4%;  Pred. No. 3.9e-127;
  Matches  963;  Conservative    0;  Mismatches  625;  Indels    6;  Gaps    1;

Qy     402 ctgcctgtcctggacccatctggggattactactactggtggctgaacacaatggtcttc 461
Db     744 CTATTTGTCTTGGACCCAGCTGGGGACTGGTACTACCGCTGGCTGTTCGTCATTGCCATG 803

Qy     462 ccagtcatgtataacctcatcatcctcgtgtgcagagcctgcttccccgacttgcagcac 521
Db     804 CCTGTGCTATACAACTGGTGTCTGCTGGTGGCCAGAGCCTGTTTCAGTGACTTGCAGAGA 863

Qy     522 ggttatctggtggcctggttggtgctggactacacgagtgacctgctatacctactagac 581
Db     864 GGCTACTTCCTGGTGTGGCTGGTGCTGGACTACTTCTCAGATGTGGTCTACATTGCGGAT 923

Qy     582 atggtggtgcgcttccacacaggattcttggaacagggcatcctggtggtggacaagggt 641
Db     924 CTTTTCATCCGATTGCGCACAGGTTTCTTGGAGCAGGGGCTGCTAGTGAAAGATCCCAAG 983

Qy     642 aggatctcgagtcgctacgttcgcacctggagtttcttcttggacctggcttccctgatg 701
Db     984 AAATTGCGAGACAACTACATCCACACCCTACAGTTCAAGCTAGATGTGGCTTCCATCATT 1043

Qy     702 cccacagatgtggtctacgtgcggctgggcccgcacacacccaccctgaggctgaaccgc 761
Db    1044 CCCACAGACCTGATCTATTTTGCTGTGGGAATCCACAACCCCGAGCTGCGCTTCAATCGT 1103

Qy     762 tttctccgcgcgccccgcctcttcgaggccttcgaccgcacagagacccgcacagcttac 821
Db    1104 CTGCTACATTTTGCCCGCATGTTTGAGTTCTTTGATCGCACTGAAACACGTACCAGTTAC 1163

Qy     822 ccaaatgcctttcgcattgccaagctgatgctttacatttttgtcgtcatccattggaac 881
Db    1164 CCCAACATCTTCCGCATCAGCAACCTAGTCCTCTACATTTTGGTCATCATCCACTGGAAT 1223

Qy     882 agctgcctatactttgccctatcccggtacctgggcttcgggcgtgacgcatgggtgtac 941
Db    1224 GCCTGTATCTACTATGCAATCTCCAAGTCCATTGGCTTTGGGGTTGATACCTGGGTTTAC 1283

Qy     942 ccggaccccgcgcagcctggctttgagcgcctgcggcgccagtacctctatagcttttac 1001
Db    1284 CCCAATATCACTGACCCTGAATATGGATACCTAGCTAGGGAGTATATCTATTGTCTTTAC 1343

Qy    1002 ttctccacgctgatactgactacagtgggcgatacaccgccgccagccagggaagaagag 1061
Db    1344 TGGTCCACACTGACCCTCACCACCATTGGGGAGACACCACCCCCTGTAAAAGATGAGGAG 1403

Qy    1062 tacctcttcatggtgggcgacttcctgctggccgtcatgggtttcgccaccatcatgggt 1121
Db    1404 TACCTATTTGTTATCTTTGACTTCCTGATTGGTGTCCTCATCTTTGCCACCATCGTGGGA 1463

Qy    1122 agcatgagctctgtcatctacaacatgaacactgcagatgcggctttctacccagatcat 1181
Db    1464 AATGTGGGCTCTATGATTTCCAATATGAATGCCACCAGGGCAGAATTCCAGGCCAAAATT 1523

Qy    1182 gcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggagcggcgagtt 1241
```

```
Db   1524 GATGCTGTCAAGCACTACATGCAGTTCCGCAAAGTCAGCAAGGAAATGGAAGCCAAGGTC 1583

Qy   1242 attgactggtatcagcacctgcagatcaacaagaagatgaccaacgaggtagccatctta 1301
Db   1584 ATTAAGTGGTTTGACTACCTGTGGACCAATAAGAAGACTGTAGATGAGCGGGAAGTCCTC 1643

Qy   1302 cagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctgtccactctg 1361
Db   1644 AAGAACCTACCAGCCAAGCTCAGGGCTGAGATAGCTATAAATGTTCACCTGTCCACACTC 1703

Qy   1362 agccgggtgcagatctttcagaactgtgaggccagcctgctggaggagctggtgctgaag 1421
Db   1704 AAGAAAGTACGCATCTTCCAGGACTGTGAGGCTGGCCTGCTGGTGGAGCTGGTGCTGAAG 1763

Qy   1422 ctgcagccccagacctactcaccaggtgaatatgtatgccgcaaaggagacattggccaa 1481
Db   1764 CTCCGTCCTCAGGTCTTCAGCCCGGGGGATTACATTTGCCGCAAGGGGGACATTGGCAAG 1823

Qy   1482 gagatgtacatcatccgagagggtcaactggccgtggtggcagatgatggtatcacacag 1541
Db   1824 GAGATGTACATAATCAAGGAAGGCAAGTTGGCAGTGGTGGCTGATGATGGTGTGACTCAG 1883

Qy   1542 tatgctgtgctcggtgcagggctctactttggggagatcagcatcatcaacatcaaaggg 1601
Db   1884 TATGCCCTGCTCTCGGCTGGGAGTTGCTTTGGCGAGATTAGCATCCTTAACATTAAAGGC 1943

Qy   1602 aacatgtctgggaaccgccgcacagccaacatcaagagcctaggttattcagacctattc 1661
Db   1944 AGCAAAATGGGCAATCGACGCACAGCCAATATTCGGAGTCTGGGCTACTCAGATCTCTTC 2003

Qy   1662 tgcctgagcaaggaggacctgcgggaggtgctgagcgagtatccacaagcacagaccatc 1721
Db   2004 TGCTTGTCCAAGGATGACCTCATGGAAGCTGTGACTGAGTATCCTGACGCCAAGAAGGTC 2063

Qy   1722 atggaggagaaaggacgtgagatcctgctgaaaatgaacaagttggacgtgaatgctgag 1781
Db   2064 TTGGAGGAGAGGGGTAGGGAGATCTTAATGAAGGAGGGGCTGCTGGATGAGAATGAGGTG 2123

Qy   1782 gcagctgagatcgccctgcaggaggccacagagtcccggctacgaggcctagaccagcag 1841
Db   2124 GCAGC------CAGCATGGAGGTAGATGTGCAAGAGAAGCTAAAGCAGCTGGAGACCAAC 2177

Qy   1842 ctggatgatctacagaccaagtttgctcgcctcctggctgagctggagtccagcgcactt 1901
Db   2178 ATGGAGACCTTGTACACTCGCTTTGGACGCCTGCTGGCAGAGTACACGGGGGCCCAGCAG 2237

Qy   1902 aagattgcttaccgcattgaacggctggagtggcagactcgagagtggccaatgcccgag 1961
Db   2238 AAACTCAAGCAGCGCATCACGGTTCTGGAGGTCAAGATGAAACAGAACACTGAGGATGAT 2297

Qy   1962 gacctggctgaggctgatgacgagggtgagcctg 1995
Db   2298 TATTTGTCAGATGGAATGAACAGTCCTGAGCCAG 2331
```

Search completed: June 20, 2002, 14:30:23
Job time: 19286 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2002, 10:53:22 ; Search time 668.44 Seconds
(without alignments)
5928.189 Million cell updates/sec

Title: US-09-927-267-2
Perfect score: 2308
Sequence: 1 agaggggaggaggaaaacag.....agactgtgcattccatctaa 2308

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2101.8	91.1	2551	24	ABA04603	MOL10b coding sequ
2	1785.4	77.4	1835	24	ABA04602	MOL10a coding sequ
3	578.6	25.1	1995	22	AAD09558	Human transporter
4	528.6	22.9	995	22	AAH98838	Rat EST-derived co
5	405.2	17.6	2909	23	ABL13353	Drosophila melanog
6	279.2	12.1	2109	23	ABL18085	Drosophila melanog
7	222.2	9.6	5133	23	ABL14931	Drosophila melanog
8	191.2	8.3	8585	23	ABL13352	Drosophila melanog
9	169.2	7.3	1125	23	AAS91258	DNA encoding novel
10	163.6	7.1	1281	23	ABL06311	Drosophila melanog
c 11	141	6.1	17150	23	ABL18084	Drosophila melanog
12	137	5.9	3275	23	ABL19457	Drosophila melanog
13	124.8	5.4	3811	23	AAS91658	DNA encoding novel
14	118.4	5.1	1146	23	AAS79945	DNA encoding novel
15	118.4	5.1	2421	23	AAS77668	DNA encoding novel
c 16	91.8	4.0	5692	23	ABL19456	Drosophila melanog
17	89.2	3.9	4751	20	AAZ09496	Human heart tissue
18	89.2	3.9	5065	22	AAH48730	Human HCN4 cDNA.
19	89.2	3.9	5499	22	ABA09197	Human cation chann
20	79	3.4	2757	24	AAD25729	Human cyclic nucle
21	78.4	3.4	1307	20	AAX25488	Human cyclic nucle
c 22	76.2	3.3	12770	23	ABL14930	Drosophila melanog
23	72.4	3.1	1820	20	AAZ09488	Bovine retina Ih i
24	70.8	3.1	1342	20	AAZ09482	Human Ih ion chann
25	70.2	3.0	1507	20	AAX84446	Mouse brain CNG-3
26	69.2	3.0	3372	22	AAH48729	Human HCN2 cDNA.
27	65.8	2.9	1790	20	AAX84445	Human brain CNG-2
28	65	2.8	2886	20	AAZ09491	Human Ih ion chann
29	62.6	2.7	3431	22	AAH98302	Human EST-derived
c 30	62.4	2.7	5645	23	ABL06310	Drosophila melanog
31	60.4	2.6	2325	21	AAC66779	Human hyperpolaris
32	59.8	2.6	3479	22	AAH21451	Human HERG1 DNA.
33	59.8	2.6	3480	21	AAA07601	Long QT syndrome a
34	59.8	2.6	3950	21	AAA07602	Long QT syndrome a
35	59.8	2.6	4070	22	AAI66257	Human potassium ch
36	59.8	2.6	4070	22	AAF61965	Human eag-related
37	56	2.4	1584	20	AAX84444	Mouse brain CNG-2
38	56	2.4	3102	22	AAH48731	Murine HCN2 cDNA.
39	55.2	2.4	3438	21	AAA62050	Hydrophobic domain
40	53.6	2.3	3355	21	AAZ50452	Monkey potassium c
41	52.2	2.3	794	20	AAZ09494	Rat olfactory Ih i
42	52	2.3	3249	21	AAZ50119	Human Elk voltage
43	52	2.3	3252	21	AAA14893	DNA encoding a her
44	52	2.3	3323	20	AAX84910	Human brain specif
45	52	2.3	3829	21	AAZ87712	Human ESK1 (hESK1)

ALIGNMENTS

RESULT 1
ABA04603
ID ABA04603 standard; cDNA; 2551 BP.
XX
AC ABA04603;
XX
DT 21-FEB-2002 (first entry)
XX
DE MOL10b coding sequence.
XX
KW MOL; G-coupled protein-receptor; cardiomyopathy; atherosclerosis;
KW cell signal processing; metabolic disorder; diabetes; cancer; human;
KW neurodegenerative disorder; immune disorder; cardiac disorder;
KW lung disease; autoimmune disease; developmental disorder; antidiabetic;
KW Cytostatic; Neuroprotective; Antiatherosclerotic; Immunosuppressive;
KW Gene therapy; Vaccine; antiinflammatory; MOL10b; chromosome 11; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 779..2506
FT /*tag= a
FT /product= "MOL 10b"
XX
PN WO200181578-A2.
XX
PD 01-NOV-2001.
XX
PF 26-APR-2001; 2001WO-US13578.
XX
PR 26-APR-2000; 2000US-200158P.
PR 28-APR-2000; 2000US-200613P.

PR 28-APR-2000; 2000US-200780P.
PR 01-MAY-2000; 2000US-201006P.
PR 01-MAY-2000; 2000US-201007P.
PR 01-MAY-2000; 2000US-201236P.
PR 01-MAY-2000; 2000US-201238P.
PR 02-MAY-2000; 2000US-201186P.
PR 03-MAY-2000; 2000US-201474P.
PR 03-MAY-2000; 2000US-201508P.
PR 25-JUL-2000; 2000US-220591P.
PR 15-SEP-2000; 2000US-232678P.
PR 22-JAN-2001; 2001US-263217P.
PR 30-JAN-2001; 2001US-265160P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Vernet CAM, Fernandes ER, Gerlach V, Shimkets RA, Malyankar UM;
PI Boldog FL, Zerhusen BD, Spytek KA, Majumder K, Tchernev VT;
PI Padigaru M, Patturajan M, Burgess CE, Gangolli EA, Smithson G;
PI Rastelli L, Macdougall JR, Taupier RJ, Grosse WM, Szekeres ES;
PI Alsobrook JP;
XX
DR WPI; 2002-049278/06.
DR P-PSDB; AAM47673.
XX
PT Novel G-protein coupled receptor-related polypeptides and
PT polynucleotides for diagnosing, preventing and treating cardiomyopathy,
PT atherosclerosis, disorders related to cell signal processing and for
PT identifying modulators -
XX
PS Claim 8; Page 78; 227pp; English.
XX
CC The present invention relates to novel G-coupled protein-receptor related
CC proteins and coding sequences (MOLX, where X is a number from 1 to 10,
CC ABA04589-ABA04603 and AAM47659-AAM47673). MOLX proteins and coding
CC sequences are useful for treating or preventing a MOLX-associated
CC disorder, such as cardiomyopathy, atherosclerosis, disorders related to
CC cell signal processing and metabolic pathway modulation, diabetes and
CC cancer. Additionally, MOLX proteins and coding sequences are useful for
CC preventing and treating a variety of disorders including metabolic
CC disorders, nutritional oedema, chronic and hereditary pancreatitis,
CC obesity, infectious disease, anorexia, neurodegenerative disorders,
CC Alzheimer's disease, Parkinson's disease, stroke, immune disorders,
CC haematopoietic disorders and various dyslipidaemias, metabolic syndrome X
CC and wasting disorders associated with chronic diseases and cancers,
CC cardiac disorders, hypertension, hypercalcaemia, cirrhosis, angiogenesis
CC and wound healing, trauma, glomerulonephritis, hyper and hypothyroidism,
CC multiple sclerosis, lung diseases including asthma, Crohn's disease,
CC scleroderma, autoimmune diseases, developmental disorders and neural tube
CC defects. The present sequence is the coding sequence for human MOL10b,
CC which maps to chromosome 11. MOL10b protein is a cyclic nucleotide-gated
CC olfactory channel-like protein.
XX
SQ Sequence 2551 BP; 646 A; 668 C; 707 G; 530 T; 0 other;

Query Match 91.1%; Score 2101.8; DB 24; Length 2551;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2103; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy      1 agaggggaggaggaaaacagagacaagactcaggcttccctctgaggcatgcacccccac 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    447 agaggggaggaggaaaacagagacaagactcaggcttccctctgaggcatgcacccccac 506

Qy     61 cttctccagggatctcattagaggtgtttagctgggcaggtgtaagcccaggccctggga 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    507 cttctccagggatctcattagaggtgtttagctgggcaggtgtaagcccaggccctggga 566

Qy    121 gacagggcagagtgctagagctagactgtctccacccccttcagtagcgctagctctggtt 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    567 gacagggcagagtgctagagctagactgtctccacccccttcagtagcgctagctctggtt 626

Qy    181 gtgttgctaagagccccaaagacaaagaagtcacagcagaagcccaacagcagcctcctt 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    627 gtgttgctaagagccccaaagacaaagaagtcacagcagaagcccaacagcagcctcctt 686

Qy    241 cagacagtcaggcactagtgcccaactccagaagtcccctacaggcagagagggtgtgga 300
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    687 caggcagtcaggcactagtgcccaactccagaagtcccctacaggcagagagggtgtgga 746

Qy    301 catctcacaccccagcaccagaccacagaaccatgagccaggacaccaaagtgaagacaa 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    747 catctcacaccccagcaccagaccacagaaccatgagccaggacaccaaagtgaagacaa 806

Qy    361 cagagtccagtcccccagccccatccaaggccaggaagttgctgcctgtcctggacccat 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    807 cagagtccagtcccccagccccatccaaggccaggaagttgctgcctgtcctggacccat 866

Qy    421 ctggggattactactactggtggctgaacacaatggtcttcccagtcatgtataacctca 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    867 ctggggattactactactggtggctgaacacaatggtcttcccagtcatgtataacctca 926

Qy    481 tcatcctcgtgtgcagagcctgcttccccgacttgcagcacggttatctggtggcctggt 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    927 tcatcctcgtgtgcagagcctgcttccccgacttgcagcacggttatctggtggcctggt 986

Qy    541 tggtgctggactacacgagtgacctgctatacctactagacatggtggtgcgcttccaca 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    987 tggtgctggactacacgagtgacctgctatacctactagacatggtggtgcgcttccaca 1046

Qy    601 caggattcttggaacagggcatcctggtggtggacaagggtaggatctcgagtcgctacg 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1047 caggattcttggaacagggcatcctggtggtggacaagggtaggatctcgagtcgctacg 1106

Qy    661 ttcgcacctggagtttcttcttggacctggcttccctgatgcccacagatgtggtctacg 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1107 ttcgcacctggagtttcttcttggacctggcttccctgatgcccacagatgtggtctacg 1166

Qy    721 tgcggctgggcccgcacacacccaccctgaggctgaaccgctttctccgcgcgccccgcc 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1167 tgcggctgggcccgcacacacccaccctgaggctgaaccgctttctccgcgcgccccgcc 1226

Qy    781 tcttcgaggccttcgaccgcacagagacccgcacagcttacccaaatgcctttcgcattg 840
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   1227 tcttcgaggccttcgaccgcgcagagacccgcacagcttacccaaatgcctttcgcattg 1286

Qy    841 ccaagctgatgctttacatttttgtcgtcatccattggaacagctgcctatactttgccc 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1287 ccaagctgatgctttacatttttgtcgtcatccattggaacagctgcctatactttgccc 1346

Qy    901 tatcccggtacctgggcttcgggcgtgacgcatgggtgtacccggaccccgcgcagcctg 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1347 tatcccggtacctgggcttcgggcgtgacgcatgggtgtacccggaccccgcgcagcctg 1406

Qy    961 gctttgagcgcctgcggcgccagtacctctatagcttttacttctccacgctgatactga 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1407 gctttgagcgcctgcggcgccagtacctctatagcttttacttctccacgctgatactga 1466

Qy   1021 ctacagtgggcgatacaccgccgccagccagggaagaagagtacctcttcatggtgggcg 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1467 ctacagtgggcgatacaccgccgccagccagggaagaagagtacctcttcatggtgggcg 1526

Qy   1081 acttcctgctggccgtcatgggtttcgccaccatcatgggtagcatgagctctgtcatct 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1527 acttcctgctggccgtcatgggtttcgccaccatcatgggtagcatgagctctgtcatct 1586

Qy   1141 acaacatgaacactgcagatgcggctttctacccagatcatgcactggtgaagaagtaca 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1587 acaacatgaacactgcagatgcggctttctacccagatcatgcactggtgaagaagtaca 1646

Qy   1201 tgaagctgcagcacgtcaaccgcaagctggagcggcgagttattgactggtatcagcacc 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1647 tgaagctgcagcacgtcaaccgcaagctggagcggcgagttattgactggtatcagcacc 1706

Qy   1261 tgcagatcaacaagaagatgaccaacgaggtagccatcttacagcacttgcctgagcggc 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1707 tgcagatcaacaagaagatgaccaacgaggtagccatcttacagcacttgcctgagcggc 1766
```

```
Qy   1321 tgcgggcagaagtggctgtgtctgtgcacctgtccactctgagccgggtgcagatctttc 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1767 tgcgggcagaagtggctgtgtctgtgcacctgtccactctgagccgggtgcagatctttc 1826

Qy   1381 agaactgtgaggccagcctgctggaggagctggtgctgaagctgcagccccagacctact 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1827 agaactgtgaggccagcctgctggaggagctggtgctgaagctgcagccccagacctact 1886

Qy   1441 caccaggtgaatatgtatgccgcaaaggagacattggccaagagatgtacatcatccgag 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1887 caccaggtgaatatgtatgccgcaaaggagacattggccaagagatgtacatcatccgag 1946

Qy   1501 agggtcaactggccgtggtggcagatgatggtatcacacagtatgctgtgctcggtgcag 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1947 agggtcaactggccgtggtggcagatgatggtatcacacagtatgctgtgctcggtgcag 2006

Qy   1561 ggctctactttggggagatcagcatcatcaacatcaaagggaacatgtctgggaaccgcc 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2007 ggctctactttggggagatcagcatcatcaacatcaaagggaacatgtctgggaaccgcc 2066

Qy   1621 gcacagccaacatcaagagcctaggttattcagacctattctgcctgagcaaggaggacc 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2067 gcacagccaacatcaagagcctaggttattcagacctattctgcctgagcaaggaggacc 2126

Qy   1681 tgcgggaggtgctgagcgagtatccacaagcacagaccatcatggaggagaaaggacgtg 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2127 tgcgggaggtgctgagcgagtatccacaagcacagaccatcatggaggagaaaggacgtg 2186

Qy   1741 agatcctgctgaaaatgaacaagttggacgtgaatgctgaggcagctgagatcgccctgc 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2187 agatcctgctgaaaatgaacaagttggacgtgaatgctgaggcagctgagatcgccctgc 2246

Qy   1801 aggaggccacagagtcccggctacgaggcctagaccagcagctggatgatctacagacca 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2247 aggaggccacagagtcccggctacgaggcctagaccagcagctggatgatctacagacca 2306

Qy   1861 agtttgctcgcctcctggctgagctggagtccagcgcacttaagattgcttaccgcattg 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2307 agtttgctcgcctcctggctgagctggagtccagcgcacttaagattgcttaccgcattg 2366

Qy   1921 aacggctggagtggcagactcgagagtggccaatgcccgaggacctggctgaggctgatg 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2367 aacggctggagtggcagactcgagagtggccaatgcccgaggacctggctgaggctgatg 2426

Qy   1981 acgagggtgagcctgaggagggaacttccaaagatgaagagggcagggccagccaggagg 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2427 acgagggtgagcctgaggagggaacttccaaagatgaagagggcagggccagccaggagg 2486

Qy   2041 gacccccaggtccagagtgaccccatccccatccccaggattcccacctcctagtgaatc 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2487 gacccccaggtccagagtgaccccatccccatccccaggattcccacctcctagtgaatc 2546

Qy   2101 cagag 2105
          |||||
Db   2547 cagag 2551
```

RESULT 2
ABA04602
ID ABA04602 standard; cDNA; 1835 BP.
XX
AC ABA04602;
XX
DT 21-FEB-2002 (first entry)
XX
DE MOL10a coding sequence.
XX
KW MOL; G-coupled protein-receptor; cardiomyopathy; atherosclerosis;
KW cell signal processing; metabolic disorder; diabetes; cancer;
KW neurodegenerative disorder; immune disorder; cardiac disorder;
KW lung disease; autoimmune disease; developmental disorder; antidiabetic;
KW Cytostatic; Neuroprotective; Antiatherosclerotic; Immunosuppressive;
KW Gene therapy; Vaccine; antiinflammatory; MOL10a; ss.
XX
OS Unidentified.
XX
FH Key Location/Qualifiers
FT CDS 54..1790
FT /*tag= a
FT /product= "MOL 10a"
XX
PN WO200181578-A2.
XX
PD 01-NOV-2001.
XX
PF 26-APR-2001; 2001WO-US13578.
XX
PR 26-APR-2000; 2000US-200158P.
PR 28-APR-2000; 2000US-200613P.
PR 28-APR-2000; 2000US-200780P.
PR 01-MAY-2000; 2000US-201006P.
PR 01-MAY-2000; 2000US-201007P.
PR 01-MAY-2000; 2000US-201236P.
PR 01-MAY-2000; 2000US-201238P.
PR 02-MAY-2000; 2000US-201186P.
PR 03-MAY-2000; 2000US-201474P.
PR 03-MAY-2000; 2000US-201508P.
PR 25-JUL-2000; 2000US-220591P.
PR 15-SEP-2000; 2000US-232678P.
PR 22-JAN-2001; 2001US-263217P.
PR 30-JAN-2001; 2001US-265160P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Vernet CAM, Fernandes ER, Gerlach V, Shimkets RA, Malyankar UM;
PI Boldog FL, Zerhusen BD, Spytek KA, Majumder K, Tchernev VT;
PI Padigaru M, Patturajan M, Burgess CE, Gangolli EA, Smithson G;
PI Rastelli L, Macdougall JR, Taupier RJ, Grosse WM, Szekeres ES;
PI Alsobrook JP;
XX
DR WPI; 2002-049278/06.
DR P-PSDB; AAM47672.
XX
PT Novel G-protein coupled receptor-related polypeptides and
PT polynucleotides for diagnosing, preventing and treating cardiomyopathy,
PT atherosclerosis, disorders related to cell signal processing and for
PT identifying modulators -
XX
PS Claim 8; Page 76; 227pp; English.
XX
CC The present invention relates to novel G-coupled protein-receptor related
CC proteins and coding sequences (MOLX, where X is a number from 1 to 10,
CC ABA04589-ABA04603 and AAM47659-AAM47673). MOLX proteins and coding
CC sequences are useful for treating or preventing a MOLX-associated
CC disorder, such as cardiomyopathy, atherosclerosis, disorders related to
CC cell signal processing and metabolic pathway modulation, diabetes and
CC cancer. Additionally, MOLX proteins and coding sequences are useful for
CC preventing and treating a variety of disorders including metabolic
CC disorders, nutritional oedema, chronic and hereditary pancreatitis,
CC obesity, infectious disease, anorexia, neurodegenerative disorders,
CC Alzheimer's disease, Parkinson's disease, stroke, immune disorders,
CC haematopoietic disorders and various dyslipidaemias, metabolic syndrome X
CC and wasting disorders associated with chronic diseases and cancers,
CC cardiac disorders, hypertension, hypercalcaemia, cirrhosis, angiogenesis
CC and wound healing, trauma, glomerulonephritis, hyper and hypothyroidism,
CC multiple sclerosis, lung diseases including asthma, Crohn's disease,
CC scleroderma, autoimmune diseases, developmental disorders and neural tube
CC defects. The present sequence is the coding sequence for MOL10a.
CC MOL10a protein is a cyclic nucleotide-gated olfactory channel-like
CC protein.
XX
SQ Sequence 1835 BP; 420 A; 521 C; 521 G; 373 T; 0 other;

Query Match 77.4%; Score 1785.4; DB 24; Length 1835;

```
  Best Local Similarity   99.5%;  Pred. No. 0;
  Matches 1825;  Conservative    0;  Mismatches    1;  Indels    9;  Gaps    3;

Qy    280 tacaggcagagagggtgtggacatctcacaccccagcaccagaccacagaaccatgagcc 339
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 tacaggcagagagggtgtggacatctcacaccccagcaccagaccacagaaccatgagcc 60

Qy    340 aggacaccaaagtgaagacaacagagtccagtcccccagccccatccaaggcc---agga 396
          |||||||||||||||||||||||||||||||||||||||||||||||||||   ||||
Db     61 aggacaccaaagtgaagacaacagagtccagtcccccagccccatccaaggccaggagga 120

Qy    397 agttgctgcctgtcctggacccatctggggattactactactggtggctgaacacaatgg 456
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 agttgctgcctgtcctggacccatctggggattactactactggtggctgaacacaatgg 180

Qy    457 tcttcccagtcatgtataacctcatcatcctcgtgtgcagagcctgcttccccgacttgc 516
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    181 tcttcccagtcatgtataacctcatcatcctcgtgtgcagggcctgcttccccgacttgc 240

Qy    517 agcacggttatctggtggcctggttggtgctggactacacgagtgacctgctatacctac 576
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 agcacggttatctggtggcctggttggtgctggactacacgagtgacctgctatacctac 300

Qy    577 tagacatggtggtgcgcttccacaca---ggattcttggaacagggcatcctggtggtgg 633
          ||||||||||||||||||||||||   ||||||||||||||||||||||||||||||||
Db    301 tagacatggtggtgcgcttccacacaggtggattcttggaacagggcatcctggtggtgg 360

Qy    634 acaagggtaggatctcgagtcgctacgttcgcacctggagtttcttcttggacctggctt 693
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 acaagggtaggatctcgagtcgctacgttcgcacctggagtttcttcttggacctggctt 420

Qy    694 ccctgatgcccacagatgtggtctacgtgcggctgggcccgcacacacccaccctgaggc 753
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ccctgatgcccacagatgtggtctacgtgcggctgggcccgcacacacccaccctgaggc 480

Qy    754 tgaaccgctttctccgcgcgccccgcctcttcgaggccttcgaccgcacagagacccgca 813
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 tgaaccgctttctccgcgcgccccgcctcttcgaggccttcgaccgcacagagacccgca 540

Qy    814 cagcttacccaaatgcctttcgcattgccaagctgatgctttacatttttgtcgtcatcc 873
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 cagcttacccaaatgcctttcgcattgccaagctgatgctttacatttttgtcgtcatcc 600

Qy    874 attggaacagctgcctatactttgccctatcccggtacctgggcttcgggcgtgacgcat 933
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 attggaacagctgcctatactttgccctatcccggtacctgggcttcgggcgtgacgcat 660

Qy    934 gggtgtacccggaccccgcgcagcctggctttgagcgcctgcggcgccagtacctctata 993
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 gggtgtacccggaccccgcgcagcctggctttgagcgcctgcggcgccagtacctctata 720

Qy    994 gcttttacttctccacgctgatactgactacagtgggcgatacaccgccgccagccaggg 1053
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 gcttttacttctccacgctgatactgactacagtgggcgatacaccgccgccagccaggg 780

Qy   1054 aagaagagtacctcttcatggtgggcgacttcctgctggccgtcatgggtttcgccacca 1113
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 aagaagagtacctcttcatggtgggcgacttcctgctggccgtcatgggtttcgccacca 840

Qy   1114 tcatgggtagcatgagctctgtcatctacaacatgaacactgcagatgcggctttctacc 1173
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 tcatgggtagcatgagctctgtcatctacaacatgaacactgcagatgcggctttctacc 900

Qy   1174 cagatcatgcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggagc 1233
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 cagatcatgcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggagc 960

Qy   1234 ggcgagttattgactggtatcagcacctgcagatcaacaagaagatgaccaacgaggtag 1293
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 ggcgagttattgactggtatcagcacctgcagatcaacaagaagatgaccaacgaggtag 1020

Qy   1294 ccatcttacagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctgt 1353
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   1021 ccatcttacagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctgt 1080

Qy   1354 ccactctgagccgggtgcagatctttcagaactgtgaggccagcctgctggaggagctgg 1413
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 ccactctgagccgggtgcagatctttcagaactgtgaggccagcctgctggaggagctgg 1140

Qy   1414 tgctgaagctgcagccccagacctactcaccaggtgaatatgtatgccgcaaaggagaca 1473
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 tgctgaagctgcagccccagacctactcaccaggtgaatatgtatgccgcaaaggagaca 1200

Qy   1474 ttggccaagagatgtacatcatccgagagggtcaactggccgtggtggcagatgatggta 1533
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 ttggccaagagatgtacatcatccgagagggtcaactggccgtggtggcagatgatggta 1260

Qy   1534 tcacacagtatgctgtgctcggtgcagggctctactttggggagatcagcatcatcaaca 1593
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 tcacacagtatgctgtgctcggtgcagggctctactttggggagatcagcatcatcaaca 1320

Qy   1594 tcaaa---gggaacatgtctgggaaccgccgcacagccaacatcaagagcctaggttatt 1650
          |||||   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 tcaaaggtgggaacatgtctgggaaccgccgcacagccaacatcaagagcctaggttatt 1380

Qy   1651 cagacctattctgcctgagcaaggaggacctgcgggaggtgctgagcgagtatccacaag 1710
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 cagacctattctgcctgagcaaggaggacctgcgggaggtgctgagcgagtatccacaag 1440

Qy   1711 cacagaccatcatggaggagaaaggacgtgagatcctgctgaaaatgaacaagttggacg 1770
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 cacagaccatcatggaggagaaaggacgtgagatcctgctgaaaatgaacaagttggacg 1500

Qy   1771 tgaatgctgaggcagctgagatcgccctgcaggaggccacagagtcccggctacgaggcc 1830
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 tgaatgctgaggcagctgagatcgccctgcaggaggccacagagtcccggctacgaggcc 1560

Qy   1831 tagaccagcagctggatgatctacagaccaagtttgctcgcctcctggctgagctggagt 1890
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 tagaccagcagctggatgatctacagaccaagtttgctcgcctcctggctgagctggagt 1620

Qy   1891 ccagcgcacttaagattgcttaccgcattgaacggctggagtggcagactcgagagtggc 1950
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 ccagcgcacttaagattgcttaccgcattgaacggctggagtggcagactcgagagtggc 1680

Qy   1951 caatgcccgaggacctggctgaggctgatgacgagggtgagcctgaggagggaacttcca 2010
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 caatgcccgaggacctggctgaggctgatgacgagggtgagcctgaggagggaacttcca 1740

Qy   2011 aagatgaagagggcagggccagccaggagggaccccccaggtccagagtgaccccatcccc 2070
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 aagatgaagagggcagggccagccaggagggaccccccaggtccagagtgaccccatcccc 1800

Qy   2071 atccccaggattcccacctcctagtgaatccagag 2105
          |||||||||||||||||||||||||||||||||||
Db   1801 atccccaggattcccacctcctagtgaatccagag 1835


RESULT   3
AAD09558
ID   AAD09558 standard; cDNA; 1995 BP.
XX
AC   AAD09558;
XX
DT   10-SEP-2001  (first entry)
XX
DE   Human transporter and ion channel-7 (TRICH-7) cDNA.
XX
KW   Human; transporter and ion channel-7; TRICH-7; vaccine; cystic fibrosis;
KW   gene therapy; amyotrophic lateral sclerosis; amnesia; muscular dystrophy;
KW   hypertension; angina; neurological disorder; asthma; bipolar disorder;
KW   dementia; depression; Alzheimer's disease; epilepsy; mood; arrhythmia;
KW   Pick's disease; ischaemic cerebrovascular disease; AIDS; anxiety; stroke;
KW   Huntington's disease; Parkinson's disease; cerebral neoplasm; allergy;
KW   demyelinating disease; mental disorder; Schizophrenia; polymyositis;
KW   muscle disorder; cardiomyopathy; cataract; myocarditis; Grave's disease;
```

KW dermatomyositis; diabetes mellitus; immunological disorder; psoriasis;
KW rheumatoid arthritis; Sjogren's syndrome; systemic lupus erythematosus;
KW sickle cell anaemia; Wilson's disease; infertility; Cushing's disease;
KW scleroderma; pulmonary artery stenosis; nootropic; Addison's disease;
KW malabsorption syndrome; hypercholesterolaemia; cancer; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1995
FT /*tag= a
FT /product= "Human TRICH-7 protein"
XX
PN WO200146258-A2.
XX
PD 28-JUN-2001.
XX
PF 22-DEC-2000; 2000WO-US35095.
XX
PR 23-DEC-1999; 99US-0172000.
PR 14-JAN-2000; 2000US-0176083.
PR 21-JAN-2000; 2000US-0177332.
PR 28-JAN-2000; 2000US-0178572.
PR 02-FEB-2000; 2000US-0179758.
PR 10-FEB-2000; 2000US-0181625.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Baughn MR, Burford N, Au-Young J, Lu DAM, Yang J, Reddy R, Lal P;
PI Hillman JL, Azimzai Y, Yue H, Nguyen DB, Yao MG, Gandhi AR;
PI Tang YT, Khan FA;
XX
DR WPI; 2001-418042/44.
DR P-PSDB; AAE04894.
XX
PT Novel human transporter and ion channel proteins useful for treating
PT and preventing transport, neurological, muscle and immunological
PT disorders -
XX
PS Claim 5; Page 145; 160pp; English.
XX
CC The present sequence is transporter and ion channel-7 (TRICH-7) cDNA.
CC TRICH is used as vaccine. TRICH is useful for treating a disease or
CC condition associated with decreased expression of functional TRICH,
CC such as transport disorder including amyotrophic lateral sclerosis,
CC cystic fibrosis, Becker's muscular dystrophy, Charcot-Marie Tooth
CC disease, Duchenne muscular dystrophy, angina and hypertension,
CC neurological disorders including Alzheimer's disease, amnesia, bipolar
CC disorder, dementia, depression, epilepsy, ischaemic cerebrovascular
CC disease, stroke, cerebral neoplasms, Pick's disease, Huntington's
CC disease and Parkinson's disease, demyelinating diseases, mental disorders
CC including mood, anxiety, Schizophrenia and seasonal affective disorder,
CC muscle disorder including cardiomyopathy, myocarditis, polymyositis,
CC dermatomyositis, arrhythmias and asthma and immunological disorders
CC including AIDS, adult respiratory distress syndrome (ARDS), allergies,
CC anaemia, diabetes mellitus, rheumatoid arthritis, scleroderma, Sjogren's
CC syndrome, systemic lupus erythematosus and other diseases including
CC sickle cell anaemia, Wilson's disease, cataracts, infertility, pulmonary
CC artery stenosis, Grave's disease, Cushing's disease, Addison's disease,
CC glucose-galactose malabsorption syndrome, hypercholesterolaemia, cancers
CC psoriasis and viral, bacterial, fungal, helminthic and protozoal
CC infections. TRICH DNA is useful in gene therapy and in diagnostic
CC purposes.
XX
SQ Sequence 1995 BP; 511 A; 511 C; 532 G; 441 T; 0 other;

Query Match 25.1%; Score 578.6; DB 22; Length 1995;
Best Local Similarity 60.7%; Pred. No. 7.1e-143;
Matches 966; Conservative 0; Mismatches 619; Indels 6; Gaps 1;

```
Qy    402 ctgcctgtcctggacccatctggggattactactactggtggctgaacacaatggtcttc 461
Db    388 ctatttgtcttggacccagctggggattggtactactgctggctatttgtcattgccatg 447

Qy    462 ccagtcatgtataacctcatcatcctcgtgtgcagagcctgcttccccgacttgcagcac 521
Db    448 cccgtcctttacaactggtgcctgctggtggccagagcctgcttcagtgacctacagaaa 507

Qy    522 ggttatctggtggcctggttggtgctggactacacgagtgacctgctatacctactagac 581
Db    508 ggctactacctggtgtggctggtgctggattatgtctcagatgtggtctacattgcggac 567

Qy    582 atggtggtgcgcttccacacaggattcttggaacagggcatcctggtggtggacaagggt 641
Db    568 ctcttcatccgattgcgcacaggtttcctggagcaggggctgctggtcaaagataccaag 627

Qy    642 aggatctcgagtcgctacgttcgcacctggagtttcttcttggacctggcttccctgatg 701
Db    628 aaactgcgagacaactacatccacaccctgcagttcaagctggatgtggcttccatcatc 687

Qy    702 cccacagatgtggtctacgtgcggctgggcccgcacacacccaccctgaggctgaaccgc 761
Db    688 cccactgacctgatctattttgctgtggacatccacagccctgaggtgcgcttcaaccgc 747

Qy    762 tttctccgcgcgccccgcctcttcgaggccttcgaccgcacagagacccgcacagcttac 821
Db    748 ctgctgcactttgcccgcatgtttgagttctttgaccggacagagacacgcaccaactac 807

Qy    822 ccaaatgcctttcgcattgccaagctgatgctttacatttttgtcgtcatccattggaac 881
Db    808 cctaacatcttccgcatcagcaaccttgtcctctacatcttggtcatcatccactggaat 867

Qy    882 agctgcctatactttgccctatcccggtacctgggcttcgggcgtgacgcatgggtgtac 941
Db    868 gcctgcatctattatgccatctccaaatccataggctttggggtcgacacctgggtttac 927

Qy    942 ccggaccccgcgcagcctggctttgagcgcctgcggcgccagtacctctatagcttttac 1001
Db    928 ccaaacatcactgaccctgagtatggctacctggctagggaatacatctattgcctttac 987

Qy   1002 ttctccacgctgatactgactacagtgggcgatacaccgccgccagccagggaagaagag 1061
Db    988 tggtccacactgactctcactaccattggggagacaccacccccgtgtaaaggatgaggag 1047

Qy   1062 tacctcttcatggtgggcgacttcctgctggccgtcatgggtttcgccaccatcatgggt 1121
Db   1048 tacctatttgtcatctttgacttcctgattggcgtcctcatctttgccaccatcgtggga 1107

Qy   1122 agcatgagctctgtcatctacaacatgaacactgcagatgcggctttctacccagatcat 1181
Db   1108 aatgtgggctccatgatctccaacatgaatgccacccgggcagagttccaggctaagatc 1167

Qy   1182 gcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggagcggcgagtt 1241
Db   1168 gatgccgtgaaacactacatgcagttccgaaaggtcagcaaggggatggaagccaaggtc 1227

Qy   1242 attgactggtatcagcacctgcagatcaacaagaagatgaccaacgaggtagccatctta 1301
Db   1228 attaggtggtttgactacttgtggaccaataagaagacagtggatgagcgagaaattctc 1287

Qy   1302 cagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctgtccactctg 1361
Db   1288 aagaatctgccagccaagctcagggctgagatagccatcaatgtccacttgtccacactc 1347

Qy   1362 agccgggtgcagatctttcagaactgtgaggccagcctgctggaggagctggtgctgaag 1421
Db   1348 aagaaagtgcgcatcttccatgattgtgaggctggcctgctggtagagctggtactgaaa 1407

Qy   1422 ctgcagccccagacctactcaccaggtgaatatgtatgccgcaaaggagacattggccaa 1481
Db   1408 ctccgtcctcaggtcttcagtcctggggattacatttgccgcaaaggggacatcggcaag 1467

Qy   1482 gagatgtacatcatccgagagggtcaactggccgtggtggcagatgatggtatcacacag 1541
Db   1468 gagatgtacatcattaaggagggcaaactggcagtggtggctgatgatggtgtgactcag 1527
```

```
Qy   1542 tatgctgtgctcggtgcagggctctactttggggagatcagcatcatcaacatcaaaggg 1601
          |||||| ||||    || ||   || |||||| |||||||| ||| | ||||| || ||
Db   1528 tatgctctgctgtcggctggaagctgctttggcgagatcagtatccttaacattaagggc 1587

Qy   1602 aacatgtctgggaaccgccgcacagccaacatcaagagcctaggttattcagacctattc 1661
          |  |     || || || |||||||| || |||   ||||| || || ||||| || |||
Db   1588 agtaaaatgggcaatcgacgcacagctaatatccgcagcctgggctactcagatctcttc 1647

Qy   1662 tgcctgagcaaggaggacctgcgggaggtgctgagcgagtatccacaagcacagaccatc 1721
          ||| ||  |||||| || ||   ||| |   |||  ||||| ||  | ||  |||   ||
Db   1648 tgcttgtccaaggatgatcttatggaagctgtgactgagtaccctgatgccaagaaagtc 1707

Qy   1722 atggaggagaaaggacgtgagatcctgctgaaaatgaacaagttggacgtgaatgctgag 1781
           | || ||||  || || ||||||||  ||||   |     | |||| | ||| |  | |
Db   1708 ctagaagagaggggtcgggagatcctcatgaaggagggactgctggatgagaacgaagtg 1767

Qy   1782 gcagctgagatcgccctgcaggaggccacagagtcccggctacgaggcctagaccagcag 1841
          |||      | |  | || |||  | |    ||     |||| |    || ||     |
Db   1768 gca------accagcatggaggtcgacgtgcaggagaagctagggcagctggagaccaac 1821

Qy   1842 ctggatgatctacagaccaagtttgctcgcctcctggctgagctggagtccagcgcactt 1901
           ||||     |  | ||   |||| ||||| |||||||||     |     |   |
Db   1822 atggaaaccttgtacactcgctttggccgcctgctggctgagtacacgggggcccagcag 1881

Qy   1902 aagattgcttaccgcattgaacggctggagtggcagactcgagagtggccaatgcccgag 1961
          ||| |     | |||||   |   |||||     |||    | ||            ||
Db   1882 aagctcaagcagcgcatcacagttctggaaaccaagatgaaacagaacaatgaagatgac 1941

Qy   1962 gacctggctgaggctgatgacgagggtgagc 1992
           ||||| |||| |      ||     |||||
Db   1942 tacctgtctgatgggatgaacagccctgagc 1972
```

RESULT 4
AAH98838
ID AAH98838 standard; cDNA; 995 BP.
XX
AC AAH98838;
XX
DT 12-OCT-2001 (first entry)
XX
DE Rat EST-derived coding sequence SEQ ID NO: 695.
XX
KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KW biodiversity; gene therapy; nutrition; ss.
XX
OS Rattus norvegicus.
XX
PN WO200154477-A2.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US02687.
XX
PR 25-JAN-2000; 2000US-0491404.
PR 17-JUL-2000; 2000US-0617746.
PR 03-AUG-2000; 2000US-0631451.
PR 15-SEP-2000; 2000US-0663870.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX
DR WPI; 2001-476164/51.
DR P-PSDB; AAM24179.
XX
PT Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use -
XX
PS Claim 1; Page 634; 1275pp; English.
XX
CC The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a cDNA
CC of the invention.
XX
SQ Sequence 995 BP; 187 A; 291 C; 269 G; 248 T; 0 other;

Query Match 22.9%; Score 528.6; DB 22; Length 995;
Best Local Similarity 97.3%; Pred. No. 9.1e-130;
Matches 548; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

```
Qy    393 aggaagttgctgcctgtcctggacccatctggggattactactactggtggctgaacaca 452
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 aggaagttgctgcctgtcctggacccatctggggattactactactggtggctgaacaca 180

Qy    453 atggtcttcccagtcatgtataacctcatcatcctcgtgtgcagagcctgcttccccgac 512
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 atggtcttcccagtcatgtataacctcatcatcctcgtgtgcagagcctgcttccccgac 240

Qy    513 ttgcagcacggttatctggtggcctggttggtgctggactacacgagtgacctgctatac 572
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ttgcagcacggttatctggtggcctggttggtgctggactacacgagtgacctgctatac 300

Qy    573 ctactagacatggtggtgcgcttccacacaggattcttggaacagggcatcctggtggtg 632
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ctactagacatggtggtgcgcttccacacaggattcttggaacagggcatcctggtggtg 360

Qy    633 gacaagggtaggatctcgagtcgctacgttcgcacctggagtttcttcttggacctggct 692
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 gacaagggtaggatctcgagtcgctacgttcgcacctggagtttcttcttggacctggct 420

Qy    693 tccctgatgcccacagatgtggtctacgtgcggctgggcccgcacacacccaccctgagg 752
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 tccctgatgcccacagatgtggtctacgtgcggctgggcccgcacacacccaccctgagg 480

Qy    753 ctgaaccgctttctccgcgcgccccgcctcttcgaggccttcgaccgcacagagacccgc 812
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ctgaaccgctttctccgcgcgccccgcctcttcgaggccttcgaccgcacagagacccgc 540

Qy    813 acagcttacccaaatgcctttcgcattgccaagctgatgctttacatttttgtcgtcatc 872
          |||||||||||||||||||| |||||| |||||||||||||||||||||| |  ||||
Db    541 acagcttacccaaatgcctttgcattggcaagctgatgctttacattttggccgcatc 600

Qy    873 cattggaacagctgcctatactttgccctatcccggtacctgggcttcgggcgtg-acgc 931
          |||||||||| |||||||||||| |||||||||||||||||||| ||||||| || |
Db    601 cattggaacaactgcctatacttttccctatcccggtacctgggctttgggcgtgaaccc 660

Qy    932 atgggtgtacccggaccccgcgc 954
          |||||||||||| |  |||  ||
Db    661 atgggtgtaccccggaccccggc 683
```

RESULT 5
ABL13353
ID ABL13353 standard; cDNA; 2909 BP.
XX
AC ABL13353;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 34541.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.

XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB69250.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 34541; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2909 BP; 729 A; 726 C; 804 G; 650 T; 0 other;

Query Match 17.6%; Score 405.2; DB 23; Length 2909;
Best Local Similarity 55.1%; Pred. No. 7e-97;
Matches 856; Conservative 0; Mismatches 668; Indels 30; Gaps 2;

```
Qy    410 cctggacccatctggggattactactactggtggctgaacacaatggtcttcccagtcat 469
          | | || ||  |   | | |   | ||| | ||||||  ||   |     |  ||||  |
Db    421 cgtcgatcccaccctgcagtcgcattacagatggctggccatcgtttcactggcagtgct 480

Qy    470 gtataacctcatcatcctcgtgtgcagagcctgcttccccgacttgcagcacggttatct 529
          || ||  ||||| || | ||| |  | ||  |||| |  || |  |  |  ||  ||
Db    481 ttacaatatcatcttcgtggtgggacgtgcggtcttctgggagatcaataagagtgctcc 540

Qy    530 ggtggcctggttggtgctggactacacgagtgacctgctatacctactagacatggtggt 589
          |    ||||   ||||||||||  | | ||| |  | || || || || | | | ||
Db    541 cgccttttggtacacgctggactacctgtgcgactttatctatctgctggatacgctcgt 600

Qy    590 gcgcttccacacaggattcttggaacagggcatcctggtggtggacaagggtaggatctc 649
          | | | |||  ||||| ||||| || ||  |||| |||  |||     | || |
Db    601 ccacatgcacgagggatttttggaccaaggtctcctcgtgcgggatgcctttcggctgcg 660

Qy    650 gagtcgctacgttcgcacctggagtttcttcttggacctggcttccctgatgcccacaga 709
          | | |||  | | ||||  | | |  | | | ||| |    ||| || ||||||| ||
Db    661 ccggcactatttccacaccaagggctggtacctcgacgtcctgtccatgctgcccacgga 720

Qy    710 tgtggtctacgtgcggctgggcccgcacacacccacc---------------------- 746
            ||| |||| |  ||  |  ||| | ||   | ||
Db    721 cttggcctacatctggtggccgccggagacctgctccagtctgtatctgccgtgtcccgt 780

Qy    747 ----ctgaggctgaaccgctttctccgcgcgccccgcctcttcgaggccttcgaccgcac 802
              || | ||||| ||  | ||| |       || || |  |||   ||||| || ||
Db    781 gatagtgcgtctgaatcgtctgctcaggatcaatcgactgtgggagtggttcgatcgaac 840

Qy    803 agagacccgcacagcttacccaaatgcctttcgcattgccaagctgatgctttacatttt 862
          |||||    || |  || || |||||||| |||||   |||| || ||||   ||||||
Db    841 ggagacagcgacgggatatccgaatgccttccgcatctgcaaggtggtgctggccatttt 900

Qy    863 tgtcgtcatccattggaacagctgcctatactttgccctatcccggtacctgggcttcgg 922
          ||  | |||||  |||||    ||| | |||||||||  ||  |    |  | ||||||  |
Db    901 ggtgctgatccactggaatgcgtgcatgtactttgccataagctacgaaatcggcttcag 960

Qy    923 gcgtgacgcatgggtgtacccggaccccgcgcagcctggctttgagcgcctgcggcgcca 982
            ||| | ||||||||     | |         |  |     |    ||| ||| ||
Db    961 ttcggactcgtgggtgta---taatctgaatggaacgaggaacaatacgttgcagcggca 1017

Qy    983 gtacctctatagcttttacttctccacgctgatactgactacagtgggcgatacaccgcc 1042
          ||  |||||||  || ||||  || |||||  |  |||||  ||  | |||||  || ||  |
Db   1018 atatatctatagtttctactggtcgacgctaacgctgaccaccattggcgagactccaac 1077

Qy   1043 gccagccagggaagaagagtacctcttcatggtgggcgacttcctgctggccgtcatggg 1102
          ||| |  |  || |  ||||| || ||| |||||| ||| ||||||   | |||| |
Db   1078 gccggagaacgatgtggagtatctgttcgtggtggccgatttcctggctggcgtcctaat 1137

Qy   1103 tttcgccaccatcatgggtagcatgagctctgtcatctacaacatgaacactgcagatgc 1162
          ||||||||||  |||||| |||  ||||  | |||| |||||||||    ||    |
Db   1138 cttcgccaccattgtgggtaacattggctccatgatctccaacatgaatgtggcccgcgt 1197

Qy   1163 ggctttctacccagatcatgcactggtgaagaagtacatgaagctgcagcacgtcaaccg 1222
          ||  ||| |          |     || ||| ||||||||    | | ||  ||    |
Db   1198 ggagttccagaatcgcatggacggagtcaagcagtacatggccttccggcgagtgggtca 1257

Qy   1223 caagctggagcggcgagttattgactggtatcagcacctgcagatcaacaagaagatgac 1282
          | |||||||   || || ||   ||||| |   ||  | |   | |       |
Db   1258 cgagctggaggcccgggtgatccgctggtttgcctacacgtggtcgcagagtggtgcgct 1317

Qy   1283 caacgaggtagccatcttacagcacttgcctgagcggctgcgggcagaagtggctgtgtc 1342
            ||||||      | |        |||| ||  ||||  ||| ||  | ||| |
Db   1318 ggacgaggagcgtgtgctggccgctctgccggacaagctgaaggcggagatcgctatcca 1377

Qy   1343 tgtgcacctgtccactctgagccgggtgcagatctttcagaactgtgaggccagcctgct 1402
          |||||| ||   || |||| | |||||  ||||| ||  ||   ||  |  | || ||
Db   1378 ggtgcacatggatacgctgaagcaggtgcgcatcttccacgacacggaaccgggtctcct 1437

Qy   1403 ggaggagctggtgctgaagctgcagccccagacctactcaccaggtgaatatgtatgccg 1462
          ||||  ||||||||||||||  | |  |||  || |   || || || ||| | || ||
Db   1438 cgaggccctggtgctgaagctcaaactgcaggtcttcagtccgggcgactatatctgtcg 1497

Qy   1463 caaaggagacattggccaagagatgtacatcatccgagagggtcaactggccgtggtggc 1522
          ||| || ||| | ||| | ||||||||||||| |       ||  |  || |||| || |
Db   1498 caagggtgacgtgggcaaggagatgtacatcgtgaagcgtggcaagttgtccgtcgtcgg 1557

Qy   1523 agatgatggtatcacacagtatgctgtgctcggtgcagggctctactttggggagatcag 1582
          |||||||| |||||    |   ||     | || || ||        || || ||| |
Db   1558 tgatgatggcatcaccgtgctggccaccttgggcgccggttcggtgttcggcgaggtttc 1617

Qy   1583 catcatcaacatcaaagggaacatgtctgggaaccgccgcacagccaacatcaagagcct 1642
          | |  || | ||    || ||   | | || || || ||||| |||||  | |    |||
Db   1618 cgtactcgagattgcgggcaatcggacgggcaatcggcgcaccgccaatgtgagatccct 1677

Qy   1643 aggttattcagacctattctgcctgagcaaggaggacctgcgggaggtgctgagcgagta 1702
          || || || || || ||||| |||   ||  |||  || |||||   |||  ||| ||
Db   1678 gggctactcggatctcttctgtctggcgaaacgggatttgtgggagaccctgtccgacta 1737

Qy   1703 tccacaagcacagaccatcatggaggagaaaggacgtgagatcctgctgaaaatgaacaa 1762
          |||  | ||     |||   |     ||  |||| |  || | |||| ||      |
Db   1738 tccggaggccagatccacgctcactcagcgaggatgccagctgctgcgcaaggatggctt 1797

Qy   1763 gttggacgtgaatgctgaggcagctgagatcgccctgcaggaggccacagagtcccggct 1822
            |||||| | |       || | |          | || ||   ||  |||  | |  |
Db   1798 actggacgagcaaatattcgctgattcacaaagggtacacgacagcatcgagggcggcat 1857

Qy   1823 acgaggcctagaccagcagctggatgatctacagaccaagtttgctcgcctcctggctga 1882
            |   || || |   | ||||  | ||  | |   | | || ||||||||||||||
Db   1858 cgaaaagctggagctctcggtggagaacctcaacatgcggctggcccgcctcctggctga 1917

Qy   1883 gctggagtccagcgcacttaagattgcttaccgcattgaacggctggagtggca 1936
          |     | | |||     |||||    | ||| | |    |||||||  | |
Db   1918 gtacacggcaagccaggccaagatcaagcagcgcctcgcgaagctggagatgaa 1971
```

```
RESULT   6
ABL18085
ID   ABL18085 standard; DNA; 2109 BP.
XX
AC   ABL18085;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 5728.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 5728; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2109 BP; 485 A; 589 C; 605 G; 430 T; 0 other;

  Query Match             12.1%;  Score 279.2;  DB 23;  Length 2109;
  Best Local Similarity   52.6%;  Pred. No. 1.4e-63;
  Matches  689;  Conservative    0;  Mismatches  603;  Indels   18;  Gaps    3;

Qy    454 tggtcttcccagtcatgtataacctcatcatcctcgtgtgcagagcctgcttccccgact 513
          ||  | |  |  || |||| ||| ||    |  |  | | | |   | |||||  ||
Db      8 tgtccatggccttcctgtacaacttctgggtgataatctaccgcttcgccttccaggaga 67

Qy    514 tgcagcacggttatctggtggcctggttggtgctggactacacgagtgacctgctatacc 573
          |  | |  |    | |     |||||    |||||||||  |   ||| | || || |
Db     68 tcaatcggcgaacgatcgccatttggttctgtctggactacctgtccgacttcctgtatc 127

Qy    574 tactagacatggtggtgcgcttccacacaggattcttggaacagggcatcctggtggtgg 633
          |  | || ||  || | | ||| | ||| |||| | ||||  | ||  | |||  |  ||
Db    128 tgatcgatatactgttccactttcgcaccggatacctggaggacggggtgctgcagacgg 187

Qy    634 acaagggtaggatctcgagtcgctacgttcgcacctggagtttcttcttggacctggctt 693
          ||      | | |     |  | |||| |    | |    ||  |||| | | |||       |
Db    188 acgctttgaagctgcgcacccactacatgaactcgacgatcttctacatcgactgcctgt 247

Qy    694 ccctgatgcccacagatgtggtctacgtgcggctgggcccgcacacacccaccctgaggc 753
          |||| ||||    ||  |  ||||  ||    | |||    ||    | | ||||  |
Db    248 gcctgctgccgctggactttctctatttgtccatcggctttaac---tcgatcctgcgca 304

Qy    754 tgaaccgctttctccgcgcgccccgcctcttcgaggccttcgaccgcacagagacccgca 813
              |||  |  |          |||  |||  |    | | ||  ||||| |||   | ||
Db    305 gcttccggctggtgaagatctaccggttctgggccttcatggatcgcaccgagcggcaca 364

Qy    814 cagcttacccaaatgcctttcgcattgccaagctgatgctttacatttttgtcgtcatcc 873
          |    ||||| ||    || || |   |    || || |  ||| |  ||||  |  |||
Db    365 ccaactacccgaacctgttccggagcacggccctcatccactacctgcttgtgatattcc 424

Qy    874 attggaacagctgcctatactttgccctatcccggtacctgggcttcgggcgtgacgcat 933
          ||||||||  | || || |||     | ||  |  | ||   || |||||      |   |
Db    425 attggaacggttgtctctaccacatcatacacaagaacaatgggttcggatcacgcaact 484

Qy    934 gggtgtacccggaccccgcgcagcctggctttgagcgcctgcggcgccagtacctctata 993
          ||||  ||||  ||| | | |  | | | | | |              |||||  ||  | |
Db    485 gggtctaccacgactcggagtcggcggacgtggtcaag---------cagtatctgcaga 535

Qy    994 gcttttacttctccacgctgatactgactacagtgggcgatacaccgccgccagccaggg 1053
          |||  ||||  | ||| |||   ||||| ||  | || |||   ||   |||   |
Db    536 gctactactggtgcaccctggccctgaccaccatcggggatcttcccaagccgcgctcca 595

Qy   1054 aagaagagtacctcttcatggtgggcgacttcctgctggccgtcatgggtttcgccacca 1113
          | |  |||||| |||| | |      |  | ||  | | | | |||   ||||||||
Db    596 agggcgagtacgtctttgttatcctgcagctgctctttggcctgatgctcttcgccacgg 655

Qy   1114 tcatgggtagcatgagctctgtcatctacaacatgaacactgcagatgcggctttctacc 1173
          |  ||||   | ||  |    |  |        |||   | ||      ||  ||  |
Db    656 tactggggcacgtggccaacattgtgacgtcagtgagtgcagcgcgcaaggagtttcaag 715

Qy   1174 cagatcatgcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggagc 1233
          |  | |  |      |||||||  |||||||  | |||  |  ||        | ||| ||
Db    716 ccaagctggatggcgtgaagacgtacatgcggatgcgacgtgtgccgaatcatctgcagg 775

Qy   1234 ggcgagttattgactggtatcagcacctgcagatcaacaagaagatgaccaacgaggtag 1293
          |   || ||  | ||||   |  |||||  | | |   | ||      |  ||||||
Db    776 tgaaggtcatcaaatggttcgattacctgtggcttacgcaaaaatgctcggacgaggagc 835

Qy   1294 ccatcttacagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctgt 1353
          |  | |        | |||||     |   ||| ||| | ||  |    || ||  |
Db    836 gcgccgtgtcctgtcttcctgataaattaaaggctgaaatagcaattaacgtccatttag 895

Qy   1354 ccactctgagccgggtgcagatctttcagaactgtgaggccagcctgctggaggagctgg 1413
            || ||  |||||| |||| || || |||   |||||| || | ||    || ||||
Db    896 atacgctcaagcgggtggagattttccaaaacacagaggccggcttcctatgcgaactgg 955

Qy   1414 tgctgaagctgcagccccagacctactcaccaggtgaatatgtatgccgcaaaggagaca 1473
          |||||   |||  ||||  |  || ||| || || || ||  | ||| | || || ||
Db    956 tgctgcgcctgaggcccgtgctcttctcgcccggcgactacatctgcagaaagggcgagg 1015

Qy   1474 ttggccaagagatgtacatcatccgagagggtcaactggccgtggtggcagatgatggta 1533
          | ||| | ||||||||||  |       || | | ||   |||||||| ||  |||| |
Db    1016 tgggcaaggagatgtacattgtgaaccgaggacgattgcaggtggtggccgacaatggaa 1075

Qy   1534 tcacacagtatgctgtgctcggtgcagggctctactttggggagatcagcatcatcaaca 1593
            ||   |   ||    ||    || ||   ||| ||||| ||||| || ||  |||| |
Db    1076 agacggtgatggcctccctgaaggctggttcctattttggcgagattagtatactcaata 1135

Qy   1594 tcaaagggaacatgtctgggaaccgccgcacagccaacatcaagagcctaggttattcag 1653
          |       |  ||   | || ||||| ||||||||| | |       | || ||    |
Db    1136 t------gggcaccgcaggcaaccgacgcacagccagcgttcgctcagtgggatacagcg 1189

Qy   1654 acctattctgcctgagcaaggaggacctgcgggaggtgctgagcgagtatccacaagcac 1713
          |||| |||  |||||||||| ||||| || |||| |||||||  ||||||||    || |
Db    1190 acctcttcgtcctgagcaagaaggacatgtgggacgtgctgaaggagtatccggcggcgc 1249

Qy   1714 agaccatcatggaggagaaaggacgtgagatcctgctgaaaatgaacaag 1763
                |||||  || ||      ||    ||  ||||   || |||
Db    1250 gtgttcgtctggagtcgatagccgtcaagcgattggagaaatacaagaag 1299

RESULT   7
```

```
ABL14931
ID   ABL14931 standard; cDNA; 5133 BP.
XX
AC   ABL14931;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 39275.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB70828.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 39275; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 5133 BP; 1201 A; 1449 C; 1501 G; 982 T; 0 other;
```

```
  Query Match             9.6%;  Score 222.2;  DB 23;  Length 5133;
  Best Local Similarity   50.7%;  Pred. No. 2.6e-48;
  Matches  713;  Conservative    0;  Mismatches  653;  Indels   39;  Gaps    6;

Qy    415 acccatctggggattactactactggtggctgaacacaatggtcttcccagtcatgtata 474
          | |||  ||   |||  |||| ||  ||| | |  |   |     |    ||  | ||||
Db      5 atccagatgaaaatttttacttctattggttaatgatgttaactgtatgtgttctatata 64

Qy    475 acctcatcatcctcgtgtgcagagcctgcttccccgacttgcagcacggttatctggtgg 534
          | ||    | |||  |    ||     |||| || ||  |||||||   |   |    |
Db     65 atctatggaccctaattgtgaggcagagctttcctgaactgcagcaatctgtgcccacgt 124

Qy    535 cctggttggtgctggactacacgagtgacctgctatacctactagacatggtggtgcgct 594
            ||| |  |    || |  | ||  ||  |  |||   | ||||| ||  | || |  |
Db    125 tttggctcatctgcgattcgatgacagatgttgtatttatcctagatataatagttcaat 184

Qy    595 tccacacaggattcttggaacagggcatcctggtggtggacaagggtaggatctcgagtc 654
          | | |||||| |   |||| ||||||  |  ||||    ||| |  | | | |  |  |||
Db    185 tacgcacaggctatctggagcagggcctaatggtatacgacgacaggaagctggcctgtc 244

Qy    655 gctacgttcgcacctggagtttcttcttggacctggcttccctgatgcccacagatgtgg 714
           |||||||| | |  |    ||  |||| ||  ||    | ||||| ||    ||| ||

Db    245 actacgttcactcgcgcgactttatcttcgatatgatagcgctgataccattggatttgc 304

Qy    715 tctacgtgcggctgggcccgcacacacccaccctgaggctgaaccgctttctccgcgcgc 774
          |  |  |   | |||||  | ||  |            |||    || |||   ||||
Db    305 tgcagctcaagatgggcacacatcc-----------gctcttgcgttttacgcgcgttt 352

Qy    775 cccgcctcttcgaggccttcgaccgcacagagacccgcacagcttacccaaatgcctttc 834
            ||     |       |   ||  |  ||| |   | ||||  ||  ||||||   |  |
Db    353 atcgatctgtgagattttattacatcgtagaaagtagaacagtttggccaaatttatggc 412

Qy    835 gcattgccaagctgatgctttacatttttgtcgtcatccattggaacagctgcctatact 894
          |  |||  || || || | |  | | ||  |  |    || |||  | | ||  | || |
Db    413 gtgttgttaacctaattcatatcctgttaatactggcacactggttcggttgtttctatt 472

Qy    895 ttgccctatcccggtacctgggcttcgggcgtgacgcatgggtgtacccggaccccgcgc 954
          ||   || |||  |     ||||||   | | |||   |||||||| ||| | |   |
Db    473 ttttactctccgaggcggagggctttcagggcgac---tgggtgtatccgtatcgacccg 529

Qy    955 agcctggctttgagcgcctgcggcgccagtacctctatagcttttacttctccacgctga 1014
            |  | || ||    |||   |||  |||| ||    ||| | ||||  || || ||||
Db    530 gg---gactatgccaccctaacgcgaaagtatctgggcagcctgtactggtcgaccctga 586

Qy   1015 tactgactacagtgggcgatacaccgccgccagccagggaagaagagtacctcttcatgg 1074
           |||||| ||  | || ||    ||  | || |  | | | | ||| ||  | || | |
Db    587 cactgacgaccattggggacctgcccacacccgaaacgaatgcagaatatatttttacga 646

Qy   1075 tgggcgacttcctgctggccgtcatgggtttcgccaccatcatgggtagcatgagctctg 1134
          | |    ||  || | |  ||  |    ||||||||||  ||||     || ||  ||
Db    647 tcgttagctatttgattggtgtttttatcttcgccaccattgtgggccaagtgggcaatg 706

Qy   1135 tcatctacaacatgaacactgcagatgcggctttctacccagatcatgcactggtgaaga 1194
          | ||   |||  ||  |           ||  ||| | |    ||  |    || ||||
Db    707 tgataacgaaccgaaatgcgaatcgcctggagttcgagcgccttctggatggggccaaga 766

Qy   1195 agtacatgaagctgcagcacgtcaaccgcaagctggagcggcgagttattgactggtatc 1254
            || |||  ||  ||  | ||   | |   | || ||||||| ||  |   |||||
Db    767 cctatatgcggcaccacaaggtgcccggagggatgaagcggcgcgtgctgcgatggtacg 826

Qy   1255 agcacctgcagatcaacaagaagatgaccaacgaggtagccatc---------ttacagc 1305
          |  |      |  |  |   | |||       |  |  | ||||         |   |
Db    827 actatagctggtcccgcggcaggatacagggtggcggtgacatcaataccgcactgggcc 886

Qy   1306 acttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctgtccactctgagcc 1365
          | |||| ||   ||||    | ||| ||||   |   |||  ||||  |  |||||
Db    887 tcctgcccgacaagctgaaaaccgaattggccttacatgtcaacctgagcgttctgaaga 946

Qy   1366 gggtgcagatctttcagaactgtgaggccagcctgctggaggagctggtgctgaagctgc 1425
           ||||  || || ||  | ||  || ||    | ||  | || || ||||| ||| ||
Db    947 aggtgaccatattccaagagtgccagcccgagttccttcacgatctcgtgctcaagatga 1006

Qy   1426 agccccagacctactcaccaggtgaatatgtatgccgcaaaggagacattggccaagaga 1485
          || || | | ||   | || || || |   | ||||| || || ||  | | ||  ||||
Db   1007 aggcctacatctttacgccgggcgactccatttgccgtaagggcgaggtggcccgcgaga 1066

Qy   1486 tgtacatcatccgagagggtcaactggccgtggtggcagatgatggtatcacacagtatg 1545
          ||| ||||||    || ||    ||||  ||| ||   ||    || |     |
Db   1067 tgttcatcattgccgatggcattctggaggtgctgagcgagacgggcaaggtcctaacca 1126

Qy   1546 ctgtgctcggtgcagggctctactttggggagatcagcatcatcaacatcaaagggaaca 1605
          |  ||      || ||    |  ||||| || ||| |||||  ||||  |  | |||
Db   1127 ccatg---aaggctggcgatttttttggcgaaatcggcatcctcaatctggacggg---- 1179

Qy   1606 tgtctgggaaccgccgcacagccaacatcaagagcctaggttattcagacctattctgcc 1665
                |||   ||||||||  |  |      | | || || || || || |||| |
Db   1180 -----cttaacaaacgcacagcggatgtgcgctccgtgggctactcggagctcttctctc 1234

Qy   1666 tgagcaaggaggacctgcgggaggtgctgagcgagtatccacaagcacagaccatcatgg 1725
          ||  |  ||||  | | || || ||  |||  || || ||  | || |||   ||  ||
Db   1235 tgtcccgcgaggatgtcctggcggccatgaaggactaccccgatgcccaggagattctgc 1294

Qy   1726 aggagaaaggacgtgagatcctgctgaaaatgaacaagttggacgtgaatgctgaggcag 1785
          ||      || ||  ||   ||  || | ||       || ||   ||    |    |
Db   1295 agaccctgggccgcaagcggctaatggaggtgcgttgcgtgaacaaaaagtacgccaagg 1354
```

```
Qy   1786 ctgagatcgccctgcaggaggccac 1810
          |  |||  | |  | ||| |||| |
Db   1355 cgcagagtgacaaggaggcggccgc 1379


RESULT   8
ABL13352
ID   ABL13352 standard; cDNA; 8585 BP.
XX
AC   ABL13352;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 34538.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB69249.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 34538; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 8585 BP; 2404 A; 1971 C; 1922 G; 2288 T; 0 other;


  Query Match             8.3%;  Score 191.2;  DB 23;  Length 8585;
  Best Local Similarity   55.2%;  Pred. No. 5.4e-40;
  Matches  373;  Conservative    0;  Mismatches  303;  Indels    0;  Gaps    0;

Qy   1119 ggtagcatgagctctgtcatctacaacatgaacactgcagatgcggctttctacccagat 1178
          |||| |||  ||||  | |||| |||||||||    ||   . | ||  ||| |
Db   5569 ggtaacattggctccatgatctccaacatgaatgtggcccgcgtggagttccagaatcgc 5628

Qy   1179 catgcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggagcggcga 1238
               |    || ||| ||||||||    | | ||  ||    | | ||||||||   ||
Db   5629 atggacggagtcaagcagtacatggccttccggcgagtgggtcacgagctggaggcccgg 5688

Qy   1239 gttattgactggtatcagcacctgcagatcaacaagaagatgaccaacgaggtagccatc 1298
          || ||   ||||| |    ||  |  |    | |       |    ||||||      |
Db   5689 gtgatccgctggtttgcctacacgtggtcgcagagtggtgcgctggacgaggagcgtgtg 5748

Qy   1299 ttacagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctgtccact 1358
               |       |||| ||   ||||  ||| ||  | ||| |    ||||| ||   ||
Db   5749 ctggccgctctgccggacaagctgaaggcggagatcgctatccaggtgcacatggatacg 5808

Qy   1359 ctgagccgggtgcagatctttcagaactgtgaggccagcctgctggaggagctggtgctg 1418
          ||||  | |||||  ||||| || ||   ||  |  | || || |||| |||||||||
Db   5809 ctgaagcaggtgcgcatcttccacgacacggaaccgggtctcctcgaggccctggtgctg 5868

Qy   1419 aagctgcagccccagacctactcaccaggtgaatatgtatgccgcaaaggagacattggc 1478
          |||||  | |  |||  || |   || || || ||| | || ||||| || ||| | |||
Db   5869 aagctcaaactgcaggtcttcagtccgggcgactatatctgtcgcaagggtgacgtgggc 5928

Qy   1479 caagagatgtacatcatccgagagggtcaactggccgtggtggcagatgatggtatcaca 1538
            | ||||||||||||| |       ||  |  || |||| || |  |||||||| |||||
Db   5929 aaggagatgtacatcgtgaagcgtggcaagttgtccgtcgtcggtgatgatggcatcacc 5988

Qy   1539 cagtatgctgtgctcggtgcagggctctactttggggagatcagcatcatcaacatcaaa 1598
            |   ||      | || || ||        || || ||| |   | |  || | ||
Db   5989 gtgctggccaccttgggcgccggttcggtgttcggcgaggtttccgtactcgagattgcg 6048

Qy   1599 gggaacatgtctgggaaccgccgcacagccaacatcaagagcctaggttattcagaccta 1658
          || ||   | | || || || ||||| |||||  | |    ||| || || || || ||
Db   6049 ggcaatcggacgggcaatcggcgcaccgccaatgtgagatccctgggctactcggatctc 6108

Qy   1659 ttctgcctgagcaaggaggacctgcgggaggtgctgagcgagtatccacaagcacagacc 1718
          |||||  |||   ||   |||  || |||||   |||  ||| |||||  | ||     ||
Db   6109 ttctgtctggcgaaacgggatttgtgggagaccctgtccgactatccggaggccagatcc 6168

Qy   1719 atcatggaggagaaaggacgtgagatcctgctgaaaatgaacaagttggacgtgaatgct 1778
          |  |     ||  |||| |  || | ||||  ||      |    |||||| | |
Db   6169 acgctcactcagcgaggatgccagctgctgcgcaaggatggcttactggacgagcaaata 6228

Qy   1779 gaggcagctgagatcg 1794
             || | | || |||
Db   6229 ttcgctggtaagttcg 6244


RESULT   9
AAS91258
ID   AAS91258 standard; cDNA; 1125 BP.
XX
AC   AAS91258;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #27062.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG27071.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
```

```
XX
PS   Claim 1; SEQ ID No 27062; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1125 BP; 283 A; 276 C; 228 G; 338 T; 0 other;


  Query Match             7.3%;  Score 169.2;  DB 23;  Length 1125;
  Best Local Similarity   55.0%;  Pred. No. 1.5e-34;
  Matches  333;  Conservative    0;  Mismatches  273;  Indels    0;  Gaps    0;

Qy    604 gattcttggaacagggcatcctggtggtggacaagggtaggatctcgagtcgctacgttc 663
          | | |||  | || ||  |  ||||   | |  |   ||   ||   | |   ||
Db    515 gttacttaaaccaaggactgttggtaaagaaagaacttaaactcataaataaatataaat 574

Qy    664 gcacctggagtttcttcttggacctggcttccctgatgcccacagatgtggtctacgtgc 723
           || || |   ||     | ||  |    || |||| || || ||| || | ||  |
Db    575 ccaacttgcaatttaaacttgatgttctgtcactgataccaactgatttgctgtatttta 634

Qy    724 ggctgggcccgcacacacccaccctgaggctgaaccgctttctccgcgcgccccgcctct 783
          | | ||   | ||   ||     | ||  | ||| | ||  | ||     | ||  | |
Db    635 agttagggtggaactatccagaaattagattaaacaggttgttacggttctctcgtatgt 694

Qy    784 tcgaggccttcgaccgcacagagacccgcacagcttacccaaatgcctttcgcattgcca 843
          | |||  |||| |  | ||||| ||  | |||   || |||||   |||  | ||| |||
Db    695 ttgagttcttccagagaacagaaacaaggacaaactatccaaacatcttcaggatttcca 754

Qy    844 agctgatgctttacatttttgtcgtcatccattggaacagctgcctatactttgccctat 903
          | ||`|  | || ||  |  || | ||||| |||||    ||  | | ||   |  | |
Db    755 accttgttatgtatatcgtcatcattatccactggaatgcatgtgtgttctactctattt 814

Qy    904 cccggtacctgggcttcgggcgtgacgcatgggtgtacccggaccccgcgcagcctggct 963
          |        | || || ||   |||  ||||||| ||||| ||        | ||||  |
Db    815 ctaaagctattggatttggaaatgatacatgggtctaccctgatattaatgatcctgaat 874

Qy    964 ttgagcgcctgcggcgccagtacctctatagcttttacttctccacgctgatactgacta 1023
          |||  ||  ||    |  | ||| | || ||| ||||||  || || ||||   ||||||
Db    875 ttggccgtttggctagaaaatacgtatacagcctttactggtctacactgactttgacta 934

Qy   1024 cagtgggcgatacaccgccgccagccagggaagaagagtacctcttcatggtgggcgact 1083
          |  | || || ||||| || || |  |||||    |||||  ||||  ||||||  || |
Db    935 ccattggtgaaacacccccctcccgtgagggattctgagtatgtctttgtggtggttgatt 994

Qy   1084 tcctgctggccgtcatgggtttcgccaccatcatgggtagcatgagctctgtcatctaca 1143
          ||||  | |  ||  |   ||| || |||||| | |||| |||  | ||| | || | ||
Db    995 tcctaattggagtgttaatttttgctaccatcgttggtaacataggttctatgatttcca 1054

Qy   1144 acatgaacactgcagatgcggctttctacccagatcatgcactggtgaagaagtacatga 1203
          |||||||  | ||    || |  ||  |  ||     ||      | ||| | || |||
Db   1055 acatgaatgcagccagagcagaatttcaagcaagaattgatgctatcaagcaatatatgc 1114

Qy   1204 agctgc 1209
```

```
          |  | |
Db   1115 attttc 1120


RESULT  10
ABL06311
ID   ABL06311 standard; cDNA; 1281 BP.
XX
AC   ABL06311;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 13415.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB62208.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 13415; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1281 BP; 315 A; 345 C; 384 G; 237 T; 0 other;


  Query Match             7.1%;  Score 163.6;  DB 23;  Length 1281;
  Best Local Similarity   56.6%;  Pred. No. 4.7e-33;
  Matches  326;  Conservative    0;  Mismatches  244;  Indels    6;  Gaps    1;

Qy   1188 gtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggagcggcgagttattgac 1247
          ||||||| |||||||  | |||  |  ||        | ||| ||  |   || ||  |
Db     13 gtgaagacgtacatgcggatgcgacgtgtgccgaatcatctgcaggtgaaggtcatcaaa 72

Qy   1248 tggtatcagcacctgcagatcaacaagaagatgaccaacgaggtagccatcttacagcac 1307
          ||||   |  |||||  | | |   | ||      |  ||||||    |  | |
Db     73 tggttcgattacctgtggcttacgcaaaaatgctcggacgaggagcgcgccgtgtcctgt 132

Qy   1308 ttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctgtccactctgagccgg 1367
           | |||||     |   ||| ||| | ||  |    || ||  |    || || |  |||
Db    133 cttcctgataaattaaaggctgaaatagcaattaacgtccatttagatacgctcaagcgg 192

Qy   1368 gtgcagatctttcagaactgtgaggccagcctgctggaggagctggtgctgaagctgcag 1427
```

```
Db    193 gtggagattttccaaaacacagaggccggcttcctatgcgaactggtgctgcgcctgagg 252

Qy   1428 ccccagacctactcaccaggtgaatatgtatgccgcaaaggagacattggccaagagatg 1487
Db    253 cccgtgctcttctcgcccggcgactacatctgcagaaagggcgaggtgggcaaggagatg 312

Qy   1488 tacatcatccgagagggtcaactggccgtggtggcagatgatggtatcacacagtatgct 1547
Db    313 tacattgtgaaccgaggacgattgcaggtggtggccgacaatggaaagacggtgatggcc 372

Qy   1548 gtgctcggtgcagggctctactttggggagatcagcatcatcaacatcaaagggaacatg 1607
Db    373 tccctgaaggctggttcctattttggcgagattagtatactcaatat------gggcacc 426

Qy   1608 tctgggaaccgccgcacagccaacatcaagagcctaggttattcagacctattctgcctg 1667
Db    427 gcaggcaaccgacgcacagccagcgttcgctcagtgggatacagcgacctcttcgtcctg 486

Qy   1668 agcaaggaggacctgcgggaggtgctgagcgagtatccacaagcacagaccatcatggag 1727
Db    487 agcaagaaggacatgtgggacgtgctgaaggagtatccggcggcgcgtgttcgtctggag 546

Qy   1728 gagaaaggacgtgagatcctgctgaaaatgaacaag 1763
Db    547 tcgatagccgtcaagcgattggagaaatacaagaag 582


RESULT  11
ABL18084/c
ID   ABL18084 standard; DNA; 17150 BP.
XX
AC   ABL18084;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 5725.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 5725; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 17150 BP; 4627 A; 3695 C; 3920 G; 4908 T; 0 other;


  Query Match             6.1%;  Score 141;  DB 23;  Length 17150;
  Best Local Similarity   51.3%;  Pred. No. 1.4e-26;
  Matches  386;  Conservative    0;  Mismatches  355;  Indels   12;  Gaps    2;

Qy    375 ccagccccatccaaggccaggaagttgctgcctgtcctggacccatctggggattactac 434
Db  16222 CGAGCAGCAACGGCGGCCAGGAAACTGCACTTCGTCTTTGATCCGGCGGGGCGACTGTGC 16163

Qy    435 tactggtggctgaacacaatggtcttcccagtcatgtataacctcatcatcctcgtgtgc 494
Db  16162 TACTACTGGTCCATGGTGGTGTCCATGGCCTTCCTGTACAACTTCTGGGTGATAATCTAC 16103

Qy    495 agagcctgcttccccgacttgcagcacggttatctggtggcctggttggtgctggactac 554
Db  16102 CGCTTCGCCTTCCAGGAGATCAATCGGCGAACGATCGCCATTTGGTTCTGTCTGGACTAC 16043

Qy    555 acgagtgacctgctatacctactagacatggtggtgcgcttccacacaggattcttggaa 614
Db  16042 CTGTCCGACTTCCTGTATCTGATCGATATACTGTTCCACTTTCGCACCGGATACCTGGAG 15983

Qy    615 cagggcatcctggtggtggacaagggtaggatctcgagtcgctacgttcgcacctggagt 674
Db  15982 GACGGGGTGCTGCAGACGGACGCTTTGAAGCTGCGCACCCACTACATGAACTCGACGATC 15923

Qy    675 ttcttcttggacctggcttccctgatgcccacagatgtggtctacgtgcggctgggcccg 734
Db  15922 TTCTACATCGACTGCCTGTGCCTGCTGCCGCTGGACTTTCTCTATTTGTCCATCGGCTTT 15863

Qy    735 cacacacccaccctgaggctgaaccgctttctccgcgcgccccgcctcttcgaggccttc 794
Db  15862 AAC---TCGATCCTGCGCAGCTTCCGGCTGGTGAAGATCTACCGGTTCTGGGCCTTCATG 15806

Qy    795 gaccgcacagagacccgcacagcttacccaaatgcctttcgcattgccaagctgatgctt 854
Db  15805 GATCGCACCGAGCGGCACACCAACTACCCGAACCTGTTCCGGAGCACGGCCCTCATCCAC 15746

Qy    855 tacatttttgtcgtcatccattggaacagctgcctatactttgccctatcccggtacctg 914
Db  15745 TACCTGCTTGTGATATTCCATTGGAACGGTTGTCTCTACCACATCATACACAAGAACAAT 15686

Qy    915 ggcttcgggcgtgacgcatgggtgtacccggaccccgcgcagcctggctttgagcgcctg 974
Db  15685 GGGTTCGGATCACGCAACTGGGTCTACCACGACTCGGAGTCGGCGGACGTGGTCAAG--- 15627

Qy    975 cggcgccagtacctctatagcttttacttctccacgctgatactgactacagtgggcgat 1034
Db  15628 ------CAGTATCTGCAGAGCTACTACTGGTGCACCCTGGCCCTGACCACCATCGGGGAT 15575

Qy   1035 acaccgccgccagccagggaagaagagtacctcttcatggtgggcgacttcctgctggcc 1094
Db  15574 CTTCCCAAGCCGCGCTCCAAGGGCGAGTACGTCTTTGTTATCCTGCAGCTGCTCTTTGGC 15515

Qy   1095 gtcatgggtttcgccaccatcatgggtagcatg 1127
Db  15514 CTGATGCTCTTCGCCACGGTACTGGGGCACGTG 15482


RESULT  12
ABL19457
ID   ABL19457 standard; DNA; 3275 BP.
XX
AC   ABL19457;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 9844.
XX
```

KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 9844; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3275 BP; 843 A; 849 C; 873 G; 710 T; 0 other;

Query Match 5.9%; Score 137; DB 23; Length 3275;
Best Local Similarity 45.7%; Pred. No. 7.9e-26;
Matches 626; Conservative 0; Mismatches 710; Indels 33; Gaps 3;

```
Qy    390 gccaggaagttgctgcctgtcctggacccatctggggattactactactggtggctgaac 449
          ||||  ||      ||    ||||||  ||    ||  |   ||||  ||  |||||   |
Db   1512 gccaacaatgatgagcgaaccctggatccccagggcaagatctacatctcctggctatgc 1571

Qy    450 acaatggtcttcccagtcatgtataacctcatcatcctcgtgtgcagagcctgcttcccc 509
          |     |  |  |  | ||||||  |    |  ||  |    | ||||  || |||
Db   1572 gttgtatcgctatcgtttctctataacgcctgggtaatcccgctgcgtgcctcgtttccc 1631

Qy    510 gacttgcagcacggttatctggtggcctggttggtgctggactacacgagtgacctgcta 569
            | |   |         ||||  |||     ||||       |||| |  ||
Db   1632 tttcaaaccaaggagaacaccaatatctggctggcctgcgacttttgcgctgacatcata 1691

Qy    570 tacctactagacatggtggtgcgcttccacacag---gattcttggaacagggcatcctg 626
          || |||||  ||  ||||  |       ||    |     | | |     ||||  ||  |
Db   1692 tatctactggatgtggtcttctttaaacatcgggttatgtacctttttgagggattctgg 1751

Qy    627 gtggtggacaagggtaggatctcgagtcgctacgttcgcacctggagtttcttcttggac 686
          |||    |||||      | |   |    |||| | || |    |   ||     ||||
Db   1752 gtgaaaaacaagaacctcaccagaaagaactacatgcgaaaactgcagtttaagctggat 1811

Qy    687 ctggcttccctgatgcccacagatgtggtctacgtgcggctgggcccgcacacacccacc 746
          |||    ||||  | ||    ||  |  | ||| |    || ||  | ||  |
Db   1812 ctgctggccctttcttccgctggagctattgtactttaaactcggaacacaggcggtttgg 1871

Qy    747 ctgaggctgaaccgctttctccgcgcgccccgcctcttcgaggccttcgaccgcacagag 806
            || |  |     | ||  |        |   || |||  || |  ||   ||     ||
Db   1872 ttgcgttttccaaggttctttaagattcagagcttctgggaagtgtttcgccttttggat 1931

Qy    807 acccgcacagcttacccaaatgcctttcgcattgccaagctgatgctttacatttttgtc 866
            || | |||  ||  |     ||||  | |||||     |    |||||  |   |
Db   1932 cgagtcatatcttctccgcactttgttcgagtggccaaaaccctaacctacatgctctac 1991

Qy    867 gtcatccattggaacagctgcctatactttgccctatcccggtacctgggcttcgggcgt 926
          | |||||    |        || ||||  |||         |||| |||  | || |
Db   1992 atgatccacatcacggctgctctgtactacgcctatagtgactaccagggtctgggacag 2051

Qy    927 gacgcatgggtgtacccggaccccgcgcagcctggctttgagcgcctgcggcgccagtac 986
          |    ||||| | |   | |   | ||| ||                          ||
Db   2052 aatcgttgggtcttcagcggcaaagggcatcc----------------------ctat 2087

Qy    987 ctctatagcttttacttctccacgctgatactgactacagtgggcgatacaccgccgcca 1046
          |     |||||   ||  ||||   ||     ||| |  | ||  | |  ||   |||
Db   2088 gtaaggtgctttgcttttgccaccaagacggcaacttcgataggaaaaaaccccaagccg 2147

Qy   1047 gccagggaagaagagtacctcttcatggtgggcgacttcctgctggccgtcatgggtttc 1106
          |   | | | |||||||| ||||||||  ||  | ||  ||  ||| |||| | |  |||
Db   2148 gaacgtcagggagagtacgtcttcatgacggtggcctggctaatgggcgtctttgtgttc 2207

Qy   1107 gccaccatcatgggtagcatgagctctgtcatctacaacatgaacactgcagatgcggct 1166
          ||     | || |||   || ||   | ||||||  |     |         |    |
Db   2208 gctctgcttatcggtcagatcagggatatcatctcgacagccactcggaataagcatgag 2267

Qy   1167 ttctacccagatcatgcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaag 1226
          |    ||      | |    |     | ||||||||   |  |   |  ||   |   ||
Db   2268 tatcgccagctggaggatgagacactggagtacatgcgacgcctcaatctctcgcaggag 2327

Qy   1227 ctggagcggcgagttattgactggtatcagcacctgcagatcaacaagaagatgaccaac 1286
          || ||  | | || |     ||||  |||         |    |  ||   |     |
Db   2328 gtgcagtcgagggtcaaaatgtggttccagtttacatgggagcagcagcgcactttggat 2387

Qy   1287 gaggtagccatcttacagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtg 1346
          ||      ||| ||  |    |||||        ||    | ||  | ||  | || |||
Db   2388 gaatccaacatattggatgctttgcccatcaatttgaaaaccgatatcgccatctccgtg 2447

Qy   1347 cacctgtccactctgagccgggtgcagatctttcagaactgtgaggccagcctgctggag 1406
          ||| |    || ||   |  ||||||  |||||    | || ||||      | ||
Db   2448 cacatccaaacgctctccaaggtgcaactctttgctgattgcgaggaggctttactcaga 2507

Qy   1407 gagctggtgctgaagctgcagccccagacctactcaccaggtgaatatgtatgccgcaaa 1466
          || ||||| || || ||   | |    || |  |  ||||| || |  || |||||||||
Db   2508 gatctggttctaaaactaagggcagtcacatttttgccaggagatttcgtttgccgcaaa 2567

Qy   1467 ggagacattggccaagagatgtacatcatccgagagggtcaactggccgtggtggcagat 1526
          || ||  | |||  |||||||||| ||| |   |  ||| ||  |    ||| |||  | |
Db   2568 ggcgaagtgggcagagagatgtatatcgtgaaactgggacaggttcaagtgatgggtggt 2627

Qy   1527 gatggtatcacacagtatgctgtgctcggtgcagggctctactttggggagatcagcatc 1586
            | ||| |       |||  ||||| |   |       | |   |   | | || |
Db   2628 cctagtagcg------atgtagtgctggccactctaaccgaaggttccgtgttcggggag 2681

Qy   1587 atcaacatcaaagggaacatgtctgggaaccgccgcacagccaacatcaagagcctaggt 1646
          ||||   |    || |  |     |   | ||| | |||||  |  ||     |   ||
Db   2682 atcagtctgctgggaataaatggagccgatcgcagaacagcagatgtccgatccaagggc 2741

Qy   1647 tattcagacctattctgcctgagcaaggaggacctgcgggaggtgctgagcgagtatcca 1706
          || ||  |||| |||   |||  |||   |||  |    |||||  |   | | |||||
Db   2742 tactccaacctgttcgtgctgtccaaatcggatttaaacgaggtcattgcctactatccc 2801

Qy   1707 caagcacagaccatcatggaggagaaaggacgtgagatcctgctgaaaa 1755
            || ||| | ||  |  |  ||   |  ||| || |  |||  ||||
Db   2802 acggcccaggcgattctcaaaaagcgtgctcgtcagctaatgcgaaaaa 2850
```

RESULT 13
AAS91658
ID AAS91658 standard; cDNA; 3811 BP.
XX
AC AAS91658;
XX
DT 13-FEB-2002 (first entry)
XX

DE DNA encoding novel human diagnostic protein #27462.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG27471.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 27462; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3811 BP; 812 A; 1093 C; 1102 G; 804 T; 0 other;

Query Match 5.4%; Score 124.8; DB 23; Length 3811;
Best Local Similarity 46.4%; Pred. No. 1.4e-22;
Matches 586; Conservative 0; Mismatches 642; Indels 36; Gaps 4;

```
Qy    503 cttccccgacttgcagcacggttatctggtggcctggttggtgctggactacacgagtga 562
Db   2092 cttcccctaccagaccccggacaacatccaccactggctgctgatggattacctatgcga 2151

Qy    563 cctgctatacctactagacat---ggtggtgcgcttccacacaggattcttggaacaggg 619
Db   2152 cctcatctacttcctggacatcaccgtgttccagacacgcctgcagtttgtcagaggcgg 2211

Qy    620 catcctggtggtggacaagggtaggatctcgagtcgctacgttcgcacctggagtttctt 679
Db   2212 ggacatcattacggacaaaaaggacatgcgaaataactacctgaagtctcgccgcttcaa 2271

Qy    680 cttggacctggcttccctgatgcccacagatgtggtctacgtgcggctgggcccgcacac 739
Db   2272 gatggacctgctcagcctcctgcccttggattttctctatttgaaagtcggtgtgaac-- 2329

Qy    740 acccaccctgaggctgaaccgctttctccgcgcgccccgcctcttcgaggccttcgaccg 799
Db   2330 -cccctcctccgcctgccccgctgtttaaagtacatggccttcttcgagtttaacagccg 2388

Qy    800 cacagagacccgcacagcttacccaaatgcctttcgcattgccaagctgatgctttacat 859
Db   2389 cctggaatccatcctcagcaaagcctacgtgtacagggtcatcaggaccacagcctacct 2448

Qy    860 ttttgtcgtcatccattggaacagctgcctatactttgccctatcccggtacctgggctt 919
Db   2449 tctctacagcctgcatttgaattcctgtctttattactgggcatcggcctatcagggcct 2508

Qy    920 cgggcgtgacgcatgggtgtacccggaccccgcgcagcctggctttgagcgcctgcggcg 979
Db   2509 cgg-----------------------ctccactcactgggtttacgatggcgtgggaaa 2544

Qy    980 ccagtacctctatagcttttacttctccacgctgatactgactacagtgggcgatacacc 1039
Db   2545 cagttatattcgctgttactactttgctgtgaagaccctcatcaccatcgggggctgcc 2604

Qy   1040 gccgccagccagggaagaagagtacctcttcatggtgggcgacttcctgctggccgtcat 1099
Db   2605 tgaccccaagacactctttgaaattgtcttccagctgctgaattatttcacgggcgtctt 2664

Qy   1100 gggtttcgccaccatcatgggtagcatgagctctgtcatctacaacatgaacactgcaga 1159
Db   2665 tgctttctctgtgatgatcggacagatgagagatgtggtaggggccgccaccgcgggaca 2724

Qy   1160 tgcggctttctacccagatcatgcactggtgaagaagtacatgaagctgcagcacgtcaa 1219
Db   2725 gacctactaccgcagctgcatggacagcacggtgaagtacatgaatttctacaagatccc 2784

Qy   1220 ccgcaagctggagcggcgagttattgactggtatcagcacctgcagatcaacaagaagat 1279
Db   2785 caagtccgtgcagaaccgcgtcaagacctggtacgagtacacctggcactcgcaaggcat 2844

Qy   1280 gaccaacgaggtagccatcttacagcacttgcctgagcggctgcgggcagaagtggctgt 1339
Db   2845 gctggatgagtcagagctgatggtgcagcttccagacaagatgcggctggacctcgccat 2904

Qy   1340 gtctgtgcacctgtccactctgagccgggtgcagatctttcagaactgtgaggccagcct 1399
Db   2905 cgacgtgaactacaacatcgttagcaaagtcgcactctttcagggctgtgaccggcagat 2964

Qy   1400 gctggaggagctggtgctgaagctgcagccccagacctactcaccaggtgaatatgtatg 1459
Db   2965 gatctttgacatgctgaagaggcttcgctctgttgtctacctgcccaacgactatgtgtg 3024

Qy   1460 ccgcaaaggagacattggccaagagatgtacatcatccgagagggtcaactggccgtggt 1519
Db   3025 caagaaggggggagatcggccgtgagatgtacatcatccaggcagggcaagtgcaggtctt 3084

Qy   1520 ggcagatgatggtatcacacagtatgctgtgctcggtgcagggctctactttggggagat 1579
Db   3085 gggcggccctgatgggaaat------ctgtgctggtgacgctgaaagctggatctgtgtt 3138

Qy   1580 cagcatcatcaacatcaaagggaacatgtctgggaaccgccgcacagccaacatcaagag 1639
Db   3139 tggagaaataagcttgctggctgttgggggcgggaaccggcgcacggccaacgtggtggc 3198

Qy   1640 cctaggttattcagacctattctgcctgagcaaggaggacctgcgggaggtgctgagcga 1699
Db   3199 gcacgggtttaccaacctcttcatcctggataagaaggacctgaatgagattttggtgca 3258

Qy   1700 gtatccacaagcacagaccatcatggaggagaaaggacgtgagatcctgctgaaaatgaa 1759
Db   3259 ttatcctgagtctcagaagttactccggaagaaagccaggcgcatgctgagaagcaacaa 3318

Qy   1760 caag 1763
Db   3319 taag 3322
```

RESULT 14
AAS79945
ID AAS79945 standard; cDNA; 1146 BP.
XX
AC AAS79945;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #15749.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG15758.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 15749; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1146 BP; 357 A; 219 C; 266 G; 304 T; 0 other;

Query Match 5.1%; Score 118.4; DB 23; Length 1146;
Best Local Similarity 70.5%; Pred. No. 4.2e-21;
Matches 158; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

```
Qy   1436 ctactcaccaggtgaatatgtatgccgcaaaggagacattggccaagagatgtacatcat 1495
          ||||   || || || ||| | |||   ||||| || || || | ||||||||||| ||
Db    360 ctacagtcctggagattatatttgcaagaaaggggatatcggacgagagatgtacattat 419

Qy   1496 ccgagagggtcaactggccgtggtggcagatgatggtatcacacagtatgctgtgctcgg 1555
          |   || ||  |||| || ||||||||||||||||  |||| |||| ||  ||  |  |
Db    420 caaggaaggcaaactcgctgtggtggcagatgatggagtcactcagtttgtggtattgag 479

Qy   1556 tgcagggctctactttggggagatcagcatcatcaacatcaaagggaacatgtctgggaa 1615
          |  ||   |||||| || ||||||||||  | ||||| ||||||| ||   |||| ||
Db    480 cgatggcagctacttcggtgagatcagcattcttaacattaaagggagcaaagctggcaa 539

Qy   1616 ccgccgcacagccaacatcaagagcctaggttattcagacctat 1659 .
          ||  | || |||||| || || ||  | || || |||||||| |
Db    540 tcgaagaacggccaatattaaaagtattggctactcagacctgt 583
```

RESULT 15
AAS77668
ID AAS77668 standard; cDNA; 2421 BP.
XX
AC AAS77668;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #13472.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG13481.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 13472; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

```
SQ   Sequence 2421 BP; 715 A; 553 C; 591 G; 562 T; 0 other;


  Query Match             5.1%;  Score 118.4;  DB 23;  Length 2421;
  Best Local Similarity   70.5%;  Pred. No. 5.8e-21;
  Matches  158;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy   1436 ctactcaccaggtgaatatgtatgccgcaaaggagacattggccaagagatgtacatcat 1495
           ||||   || || || ||| | |||   ||||| || || || | ||||||||||| ||
Db    360 ctacagtcctggagattatatttgcaagaaaggggatatcggacgagagatgtacattat 419

Qy   1496 ccgagagggtcaactggccgtggtggcagatgatggtatcacacagtatgctgtgctcgg 1555
           |   || ||  |||| || ||||||||||||||||  |||| |||| ||  ||  |  |
Db    420 caaggaaggcaaactcgctgtggtggcagatgatggagtcactcagtttgtggtattgag 479

Qy   1556 tgcagggctctactttggggagatcagcatcatcaacatcaaagggaacatgtctgggaa 1615
           |  ||   |||||| || ||||||||||  | ||||| ||||||| ||   |||| ||
Db    480 cgatggcagctacttcggtgagatcagcattcttaacattaaagggagcaaagctggcaa 539

Qy   1616 ccgccgcacagccaacatcaagagcctaggttattcagacctat 1659
           ||  | || ||||| || || ||  | || || ||||||||
Db    540 tcgaagaacggccaatattaaaagtattggctactcagacctgt 583


Search completed: June 20, 2002, 14:38:56
Job time: 13534 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2002, 09:17:57 ; Search time 155.16 Seconds
(without alignments)
3653.790 Million cell updates/sec

Title: US-09-927-267-2
Perfect score: 2308
Sequence: 1 agaggggaggaggaaaacag.....agactgtgcattccatctaa 2308

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	78.4	3.4	1307	4	US-09-172-422-3	Sequence 3, Appli
	2	59.8	2.6	3480	4	US-09-226-012-1	Sequence 1, Appli
	3	59.8	2.6	3950	4	US-09-226-012-3	Sequence 3, Appli
c	4	52.8	2.3	7218	1	US-08-232-463-14	Sequence 14, Appl
	5	52.4	2.3	7218	1	US-08-232-463-14	Sequence 14, Appl
	6	52	2.3	3323	4	US-09-600-776-1	Sequence 1, Appli
	7	49.8	2.2	3064	4	US-09-600-776-5	Sequence 5, Appli
	8	49	2.1	3141	2	US-08-956-242-1	Sequence 1, Appli
	9	49	2.1	3141	3	US-09-351-215-1	Sequence 1, Appli
c	10	45.2	2.0	84495	4	US-09-797-906-3	Sequence 3, Appli
	11	44	1.9	4722	4	US-08-979-608A-14	Sequence 14, Appl
	12	42.8	1.9	818	4	US-08-205-697A-4	Sequence 4, Appli
	13	42.8	1.9	818	4	US-08-702-525-4	Sequence 4, Appli
	14	42.8	1.9	818	5	PCT-US95-02576-4	Sequence 4, Appli
	15	42.8	1.9	1570	4	US-08-205-697A-10	Sequence 10, Appl
	16	42.8	1.9	1570	4	US-08-702-525-10	Sequence 10, Appl
	17	42.8	1.9	1570	5	PCT-US95-02576-10	Sequence 10, Appl
	18	42.8	1.9	1606	4	US-08-702-525-64	Sequence 64, Appl
	19	42.8	1.9	1606	5	PCT-US95-02576-64	Sequence 64, Appl
	20	42.8	1.9	1888	4	US-08-205-697A-1	Sequence 1, Appli
	21	42.8	1.9	1888	4	US-08-702-525-1	Sequence 1, Appli
	22	42.8	1.9	1888	5	PCT-US95-02576-1	Sequence 1, Appli
	23	42.8	1.9	2516	4	US-08-205-697A-3	Sequence 3, Appli
	24	42.8	1.9	2516	4	US-08-702-525-3	Sequence 3, Appli
	25	42.8	1.9	2516	5	PCT-US95-02576-3	Sequence 3, Appli
	26	41.8	1.8	809	2	US-08-370-909-1	Sequence 1, Appli
	27	41.8	1.8	809	4	US-09-056-105-22	Sequence 22, Appl
	28	41	1.8	1995	1	US-08-425-069-3	Sequence 3, Appli
	29	41	1.8	1995	2	US-08-317-844B-3	Sequence 3, Appli
c	30	40.6	1.8	270	1	US-08-222-177A-51	Sequence 51, Appl
	31	40.4	1.8	2580	3	US-09-050-863-2	Sequence 2, Appli
	32	40.4	1.8	2580	4	US-09-359-081-2	Sequence 2, Appli
c	33	40.4	1.8	5452	2	US-09-130-114-1	Sequence 1, Appli
	34	40.4	1.8	9600	4	US-08-910-647-1	Sequence 1, Appli
	35	40.4	1.8	10596	1	US-07-884-811-15	Sequence 15, Appl
	36	40.4	1.8	10596	1	US-07-885-971-15	Sequence 15, Appl
	37	40.4	1.8	10596	1	US-08-087-783A-15	Sequence 15, Appl
	38	40.4	1.8	10596	1	US-08-194-088B-15	Sequence 15, Appl
	39	40.4	1.8	10596	2	US-08-194-087-15	Sequence 15, Appl
	40	40.4	1.8	10596	5	PCT-US93-04648-15	Sequence 15, Appl
	41	39.8	1.7	5024	4	US-09-307-143-1	Sequence 1, Appli
	42	39.8	1.7	10409	3	US-08-772-440-33	Sequence 33, Appl
	43	39	1.7	2338	1	US-08-425-069-1	Sequence 1, Appli
	44	39	1.7	2338	2	US-08-317-844B-1	Sequence 1, Appli
	45	38.6	1.7	1007	3	US-09-248-335-31	Sequence 31, Appl

ALIGNMENTS

RESULT 1
US-09-172-422-3
; Sequence 3, Application US/09172422A
; Patent No. 6300485
; GENERAL INFORMATION:
; APPLICANT: Adams, Arwen E.
; APPLICANT: Chiu, Choi Ying
; APPLICANT: Duhl, David
; APPLICANT: Gorman, Susan W.
; APPLICANT: Leng, Song
; APPLICANT: Sheffield, Val
; APPLICANT: Welch, Juliet
; TITLE OF INVENTION: MYOSIN IXA AND CYCLIC NUCLEOTIDE GATED
; TITLE OF INVENTION: CHANNEL-15 (CNGC-15) POLYNUCLEOTIDES, POLYPEPTIDES,
; TITLE OF INVENTION: COMPOSITIONS, METHODS, AND USES THEREOF
; FILE REFERENCE: 200130.442
; CURRENT APPLICATION NUMBER: US/09/172,422A
; CURRENT FILING DATE: 1998-10-14
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1307
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1307)
; OTHER INFORMATION: n = A,T,C or G
US-09-172-422-3

Query Match 3.4%; Score 78.4; DB 4; Length 1307;
Best Local Similarity 46.6%; Pred. No. 8.3e-12;
Matches 324; Conservative 0; Mismatches 366; Indels 6; Gaps 2;

```
Qy    832 ttcgcattgccaagctgatgctttacatttttgtcgtcatccattggaacagctgcctat 891
          | ||||| |  || || ||       ||  |  |  ||  ||| ||| || |||||||
Db    590 tgcgcatcgtgaacctcatcggcatgatgctcctgctctgccactgggacggctgcctgc 649

Qy    892 actttgccctatcccggtacctgggcttcgggcgtgacgcatgggtgtacccggaccccg 951
          | ||     || ||  |   | || ||||      |||   ||||||| |    ||  |
Db    650 agttcctggtacccatgctacaggacttccctgacgactgctgggtgtccatcaacaaca 709

Qy    952 cgcagcctggctttgagcgcctgcggcgccagtacctctatagcttttacttctccacgc 1011
          |  |     ||    | |   |  ||   | | |  |     |||  |   |   | |
Db    710 tggtgaacaactcctgggg---gaagcagtactcctacgcgctcttcaaggccatgagcc 766

Qy   1012 tgatactgactacagtgggcgatacaccgccgccagccagggaagaagagtacctcttca 1071
            || |||   |  |    ||     | | |||| |   |       ||   |    |||
Db    767 acatgctgtgcatcggctacgggcggcaggcgcccgtgggcatctccgacgtctggctca 826
```

```
Qy   1072 tggtgggcgacttcctgctggccgtcatgggtttcgccaccatcatgggtagcatgagct 1131
Db    827 ccatgctcagcatgatcgtgggtgccacctgctacgccatgttcattggccacgccactg 886

Qy   1132 ctgtcatctacaacatgaacactgcagatgcggctttctacccagatcatgcactggtga 1191
Db    887 ccctcatccagtccctggactcctcccggcgccagtaccaggaaaagtacaagcaggtgg 946

Qy   1192 agaagtacatgaagctgcagcacgtcaaccgcaagctggagcggcgagttattgactggt 1251
Db    947 agcagtacatgtcctttcacaagctcccgcccgacacccggcagcgcatccacgactact 1006

Qy   1252 atcagcacctgcagatcaacaagaagatgaccaacgaggtagccatcttacagcacttgc 1311
Db   1007 acgagcaccgcta---ccagggcaagatgttcgacgaggagagcaccctgggcgagctaa 1063

Qy   1312 ctgagcggctgcgggcagaagtggctgtgtctgtgcacctgtccactctgagccgggtgc 1371
Db   1064 gcgagcccctgcgggaggagatcatcaacttaactgtcggaagctggtggcctccatgc 1123

Qy   1372 agatctttcagaactgtgaggccagcctgctggaggagctggtgctgaagctgcagcccc 1431
Db   1124 cactgtttgccaatgcggaccccaacttcgtgacgtccatgctgaccaagctgcgtttcg 1183

Qy   1432 agacctactcaccaggtgaatatgtatgccgcaaaggagacattggccaagagatgtaca 1491
Db   1184 aggtcttccagcctggggactacatcatccgggaaggcaccattggcaagaagatgtact 1243

Qy   1492 tcatccgagagggtcaactggccgtggtggcagatg 1527
Db   1244 tcatccagcatggcgtggtcagcgtgctcaccaagg 1279
```

RESULT 2
US-09-226-012-1
; Sequence 1, Application US/09226012
; Patent No. 6207383
; GENERAL INFORMATION:
; APPLICANT: Keating, Mark T.
; APPLICANT: Splawski, Igor
; TITLE OF INVENTION: MUTATIONS IN AND GENOMIC STRUCTURE OF HERG - A LONG QT
; TITLE OF INVENTION: SYNDROME GENE
; FILE REFERENCE: 2323-136
; CURRENT APPLICATION NUMBER: US/09/226,012
; CURRENT FILING DATE: 1999-01-06
; EARLIER APPLICATION NUMBER: 09/122,847
; EARLIER FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3480
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3477)
US-09-226-012-1

Query Match 2.6%; Score 59.8; DB 4; Length 3480;
Best Local Similarity 45.9%; Pred. No. 1.5e-06;
Matches 205; Conservative 0; Mismatches 242; Indels 0; Gaps 0;

```
Qy   1058 agagtacctcttcatggtgggcgacttcctgctggccgtcatgggtttcgccaccatcat 1117
Db   1908 agagaagatcttctccatctgcgtcatgctcattggctccctcatgtatgctagcatctt 1967

Qy   1118 gggtagcatgagctctgtcatctacaacatgaacactgcagatgcggctttctacccaga 1177
Db   1968 cggcaacgtgtcggccatcatccagcggctgtactcgggcacagcccgctaccacacaca 2027

Qy   1178 tcatgcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggagcggcg 1237
Db   2028 gatgctgcgggtgcgggagttcatccgcttccaccagatccccaatcccctgcgccagcg 2087

Qy   1238 agttattgactggtatcagcacctgcagatcaacaagaagatgaccaacgaggtagccat 1297
Db   2088 cctcgaggagtacttccagcacgcctggtcctacaccaacggcatcgacatgaacgcggt 2147

Qy   1298 cttacagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctgtccac 1357
Db   2148 gctgaagggcttccctgagtgcctgcaggctgacatctgcctgcacctgaaccgctcact 2207

Qy   1358 tctgagccgggtgcagatctttcagaactgtgaggccagcctgctggaggagctggtgct 1417
Db   2208 gctgcagcactgcaaacccttccgaggggccaccaagggctgccttcgggccctggccat 2267

Qy   1418 gaagctgcagccccagacctactcaccaggtgaatatgtatgccgcaaaggagacattgg 1477
Db   2268 gaagttcaagaccacacatgcaccgccaggggacacactggtgcatgctggggacctgct 2327

Qy   1478 ccaagagatgtacatcatccgagaggg 1504
Db   2328 caccgccctgtacttcatctcccgggg 2354
```

RESULT 3
US-09-226-012-3
; Sequence 3, Application US/09226012
; Patent No. 6207383
; GENERAL INFORMATION:
; APPLICANT: Keating, Mark T.
; APPLICANT: Splawski, Igor
; TITLE OF INVENTION: MUTATIONS IN AND GENOMIC STRUCTURE OF HERG - A LONG QT
; TITLE OF INVENTION: SYNDROME GENE
; FILE REFERENCE: 2323-136
; CURRENT APPLICATION NUMBER: US/09/226,012
; CURRENT FILING DATE: 1999-01-06
; EARLIER APPLICATION NUMBER: 09/122,847
; EARLIER FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 3950
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (67)..(3543)
US-09-226-012-3

Query Match 2.6%; Score 59.8; DB 4; Length 3950;
Best Local Similarity 45.9%; Pred. No. 1.5e-06;
Matches 205; Conservative 0; Mismatches 242; Indels 0; Gaps 0;

```
Qy   1058 agagtacctcttcatggtgggcgacttcctgctggccgtcatgggtttcgccaccatcat 1117
Db   1974 agagaagatcttctccatctgcgtcatgctcattggctccctcatgtatgctagcatctt 2033

Qy   1118 gggtagcatgagctctgtcatctacaacatgaacactgcagatgcggctttctacccaga 1177
Db   2034 cggcaacgtgtcggccatcatccagcggctgtactcgggcacagcccgctaccacacaca 2093

Qy   1178 tcatgcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggagcggcg 1237
Db   2094 gatgctgcgggtgcgggagttcatccgcttccaccagatccccaatcccctgcgccagcg 2153

Qy   1238 agttattgactggtatcagcacctgcagatcaacaagaagatgaccaacgaggtagccat 1297
Db   2154 cctcgaggagtacttccagcacgcctggtcctacaccaacggcatcgacatgaacgcggt 2213

Qy   1298 cttacagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctgtccac 1357
Db   2214 gctgaagggcttccctgagtgcctgcaggctgacatctgcctgcacctgaaccgctcact 2273
```

```
Qy   1358 tctgagccgggtgcagatctttcagaactgtgaggccagcctgctggaggagctggtgct 1417
          |||   |       |   ||| |               ||   ||   ||  ||||   |
Db   2274 gctgcagcactgcaaacccttccgaggggccaccaagggctgccttcgggccctggccat 2333

Qy   1418 gaagctgcagccccagacctactcaccaggtgaatatgtatgccgcaaaggagacattgg 1477
          |||| |  || ||          | ||||| ||     |    |     || ||| |
Db   2334 gaagttcaagaccacacatgcaccgccaggggacacactggtgcatgctggggacctgct 2393

Qy   1478 ccaagagatgtacatcatccgagaggg 1504
          |   |   ||||| |||||     |||
Db   2394 caccgccctgtacttcatctcccgggg 2420


RESULT   4
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14


  Query Match             2.3%;  Score 52.8;  DB 1;  Length 7218;
  Best Local Similarity   5.8%;  Pred. No. 0.00016;
  Matches   24;  Conservative  218;  Mismatches  170;  Indels    0;  Gaps    0;

Qy   1694 gagcgagtatccacaagcacagaccatcatggaggagaaaggacgtgagatcctgctgaa 1753
          ::: ::: :   : ::: : :::  :  : :::::::::::::: : ::::    :  :::
Db   1431 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1372

Qy   1754 aatgaacaagttggacgtgaatgctgaggcagctgagatcgccctgcaggaggccacaga 1813
          :: ::: :::  ::: : ::: :  :::: ::  ::::  :    : ::::::  : :::
Db   1371 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1312

Qy   1814 gtcccggctacgaggcctagaccagcagctggatgatctacagaccaagtttgctcgcct 1873
          :   ::  : ::::  :::  :: ::  ::: ::   : :::  :::   :   :
Db   1311 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1252

Qy   1874 cctggctgagctggagtccagcgcacttaagattgcttaccgcattgaacggctggagtg 1933
            ::  :::  ::::   :: : :   ::::  :   :  : :  ::: ::  :::: :
Db   1251 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1192

Qy   1934 gcagactcgagagtggccaatgcccgaggacctggctgaggctgatgacgagggtgagcc 1993
          : :::   ::::: ::  :: :   :::::   ::  ::::  :: :: ::::: :::
Db   1191 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1132

Qy   1994 tgaggagggaacttccaaagatgaagagggcagggccagccaggagggacccccaggtcc 2053
           ::::::::::     ::::: :::::::: ::::  ::  ::::::::     :::
Db   1131 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1072

Qy   2054 agagtgaccccatccccatccccaggattcccacctcctagtgaatccagag 2105
          ::::   | | | | || ||  |  |  ¬|| | |||  | | ||| | | |
Db   1071 RRRRRATCGCAAGCTCCCTCGACCTGCAGCCAAGCTCGGAATTAATTCTGTG 1020


RESULT   5
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14
```

```
  Query Match             2.3%;  Score 52.4;  DB 1;  Length 7218;
  Best Local Similarity   8.2%;  Pred. No. 0.00021;
  Matches   32;  Conservative  197;  Mismatches  163;  Indels    0;  Gaps    0;

Qy     646 tctcgagtcgctacgttcgcacctggagtttcttcttggacctggcttccctgatgccca 705
           ::::   :: :: : ::: : :::    :::::::::   :::  :::::::  : :::
Db    1105 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1164

Qy     706 cagatgtggtctacgtgcggctgggcccgcacacacccaccctgaggctgaaccgctttc 765
           :   : :  ::: : : :  ::   ::: : : : ::: ::::    ::   :: :::::
Db    1165 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1224

Qy     766 tccgcgcgccccgcctcttcgaggccttcgaccgcacagagacccgcacagcttacccaa 825
           ::: : : :::: :::::::    :::::  :: : :     ::: : :  ::: :::
Db    1225 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1284

Qy     826 atgcctttcgcattgccaagctgatgctttacatttttgtcgtcatccattggaacagct 885
           : :::::: : :: ::   ::  : :::: : ::::: :: :: ::: ::    :  ::
Db    1285 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1344

Qy     886 gcctatactttgccctatcccggtacctgggcttcgggcgtgacgcatgggtgtacccgg 945
           ::: : :::: :::: ::::  : :::   ::::   : :  : : :    : : :::
Db    1345 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1404

Qy     946 accccgcgcagcctggctttgagcgcctgcggcgccagtacctctatagcttttacttct 1005
           :::: : :  :::  ::::   : ::: :|   | | | | |||||  |||| ||| ||
Db    1405 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYGTACCAAATTCTTCTATCTCTTTAACTACT 1464

Qy    1006 ccacgctgatactgactacagtgggcgataca 1037
                     | | |||||||     ||||
Db    1465 TGCATAGATAGGTAATTACAGTGATGCCTACA 1496


RESULT   6
US-09-600-776-1
; Sequence 1, Application US/09600776
; Patent No. 6326168
; GENERAL INFORMATION:
;  APPLICANT:  Yamanouchi Pharmaceutical Co., Ltd.
;  TITLE OF INVENTION:  A novel potassium channel protein
;  FILE REFERENCE:  Y9903-PCT
;  CURRENT APPLICATION NUMBER: US/09/600,776
;  CURRENT FILING DATE:  2000-07-21
;  PRIOR APPLICATION NUMBER:  JP P1998-011434
;  PRIOR FILING DATE:  1998-01-23
;  PRIOR APPLICATION NUMBER:  JP P1998-346198
;  PRIOR FILING DATE:  1998-12-04
;  NUMBER OF SEQ ID NOS:  12
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 3323
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (6)..(3257)
US-09-600-776-1


  Query Match             2.3%;  Score 52;  DB 4;  Length 3323;
  Best Local Similarity   45.4%;  Pred. No. 0.00019;
  Matches  267;  Conservative    0;  Mismatches  315;  Indels    6;  Gaps    2;

Qy     943 cggaccccgcgcagcctggctttgagcgcctgcggcgccagtacctctatagcttttact 1002
           |||  |  | || ||  |||     |  |||||  ||   ||| ||     | | ||||
Db    1318 cggggctggagctgctgggcggcccgtcgctgcgcagcgcctacatcacctccctctact 1377

Qy    1003 tctccacgctgatactgactacagtggg---cgatacaccgccgccagccagggaagaag 1059
           || |       |  || || |  |||||   ||  |    | |  |   || |||    |
Db    1378 tcgcactcagcagcctcaccagcgtgggcttcggcaacgtgtccgccaacacggacaccg 1437

Qy    1060 agtacctcttcatggtgggcgacttcctgctggccgtcatgggtttcgccaccatcatgg 1119
           || |  |||||    |  ||  | | ||  | | || | ||     |||     | | |
Db    1438 agaagatcttctccatctgcaccatgctcatcggcgccctgatgcacgcggtggtgtttg 1497

Qy    1120 gtagcatgagctctgtcatctacaacatgaacactgcagatgcggctttctacccagatc 1179
           | | | |||   |  |||||  |   |||| || |              | | ||
Db    1498 ggaacgtgacggccatcatccagcgcatgtacgcccgccgctttctgtaccacagccgca 1557

Qy    1180 atgcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggagcggcgag 1239
                 ||    | |||||     | || |   ||  |      ||  ||| |||
Db    1558 cgcgcgacctgcgcgactacatccgcatccaccgtatccccaagcccctcaagcagcgca 1617

Qy    1240 ttattgactggtatcagcacctgcagatcaacaagaagatgaccaacgaggtagccatct 1299
           |  | || |  |  |||  |      |       || ||    | | ||      |  |
Db    1618 tgctggagtacttccaggccacctgggcggtgaacaatggcatcgacaccaccgagctgc 1677

Qy    1300 tacagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctgtccactc 1359
           | |||  | | |||||   ||||||  |||||  | ||  ||    |||||  |      |
Db    1678 tgcagagcctccctgacgagctgcgcgcagacatcgccatgcacctgcacaaggaggtcc 1737

Qy    1360 tgagccgggtgcagatcttcagaactgtgaggccagcctgctggaggagctggtgctga 1419
           ||   | | |||   | |||  ||           | ||   |||  ||  |||   |||
Db    1738 tg---cagctgccactgtttgaggcggccagccgcggctgcctgcgggcactgtctctgg 1794

Qy    1420 agctgcagccccagacctactcaccaggtgaatatgtatgccgcaaaggagacattggcc 1479
             |||| ||||     || | | || || || ||  |   || | |||| ||       |
Db    1795 ccctgcggcccgccttctgcacgccgggcgagtacctcatccaccaaggcgatgccctgc 1854

Qy    1480 aagagatgtacatcatccgagagggtcaactggccgtggtggcagatg 1527
           | |   | ||| |  || |    ||     |||  ||| |    | ||
Db    1855 aggccctctactttgtctgctctggctccatggaggtgctcaagggtg 1902


RESULT   7
US-09-600-776-5
; Sequence 5, Application US/09600776
; Patent No. 6326168
; GENERAL INFORMATION:
;  APPLICANT:  Yamanouchi Pharmaceutical Co., Ltd.
;  TITLE OF INVENTION:  A novel potassium channel protein
;  FILE REFERENCE:  Y9903-PCT
;  CURRENT APPLICATION NUMBER: US/09/600,776
;  CURRENT FILING DATE:  2000-07-21
;  PRIOR APPLICATION NUMBER:  JP P1998-011434
;  PRIOR FILING DATE:  1998-01-23
;  PRIOR APPLICATION NUMBER:  JP P1998-346198
;  PRIOR FILING DATE:  1998-12-04
;  NUMBER OF SEQ ID NOS:  12
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 3064
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (4)..(3057)
US-09-600-776-5


  Query Match             2.2%;  Score 49.8;  DB 4;  Length 3064;
  Best Local Similarity   45.5%;  Pred. No. 0.00072;
  Matches  217;  Conservative    0;  Mismatches  257;  Indels    3;  Gaps    1;

Qy    1049 cagggaagaagagtacctcttcatggtgggcgacttcctgctggccgtcatgggtttcgc 1108
           ||  || |  ||| |  |||||     |  ||    | || | | || | ||     |||
Db    1347 caccgacgcggagaagatcttctccatctgcacgatgctcataggcgccctgatgcacgc 1406

Qy    1109 caccatcatgggtagcatgagctctgtcatctacaacatgaacactgcagatgcggcttt 1168
              | | || | | |||   |  ||||| |   |||| || |       ||   |
Db    1407 tgtggtgttcgggaacgtgacagccatcatccagcgcatgtactcgcgccgctcgctcta 1466

Qy    1169 ctacccagatcatgcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaagct 1228
           | ||                | ||| | | |||     |||| | | |    ||   |||
Db    1467 ccacagccgcatgaaggacctcaaggacttcatccgtgtgcaccgcctgccgcggccgct 1526
```

```
Qy   1229 ggagcggcgagttattgactggtatcagcacctgcagatcaacaagaagatgaccaacga 1288
          ||| |||  |  | || |  |  |||  |  |  |  |  ||| |     | | |||
Db   1527 caagcagcgcatgctcgaatacttccagaccacgtgggccgtcaacagcggcatcgacgc 1586

Qy   1289 ggtagccatcttacagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtgca 1348
          |   |  | |   |||| || ||   |||| | || ||  | ||| ||    || |
Db   1587 caacgagttactgcgtgacttcccagacgagctgagagctgacattgctatgcacctgaa 1646

Qy   1349 cctgtccactctgagccgggtgcagatctttcagaactgtgaggccagcctgctggagga 1408
          | |   |  |||   | |  |  | ||          |   || | |||||| ||
Db   1647 tcgggagatcctgcagctgccgttgttcggggcagcgagcaggggctgcctgcgggcc-- 1704

Qy   1409 gctggtgctgaagctgcagccccagacctactcaccaggtgaatatgtatgccgcaaagg 1468
          ||   |||| |  |  || ||  |  || | | || || || ||  | |  |||   ||
Db   1705 -ctatcgctgcacatcaagacctcgttctgcgctccgggcgagtacctgttgcgccgtgg 1763

Qy   1469 agacattggccaagagatgtacatcatccgagagggtcaactggccgtggtggcaga 1525
          ||       || |     |||    || |    ||    || |  ||| |   |||
Db   1764 ggatgccctgcaggcacattactatgtctgctccggctcgcttgaggtgctccgaga 1820


RESULT   8
US-08-956-242-1
; Sequence 1, Application US/08956242C
; Patent No. 5986081
; GENERAL INFORMATION:
;  APPLICANT: Ganetzky, Barry S.
;  APPLICANT: Titus, Steven A.
;  TITLE OF INVENTION: Polynucleotides Encoding Herg-3
;  FILE REFERENCE: 960296.94550
;  CURRENT APPLICATION NUMBER: US/08/956,242C
;  CURRENT FILING DATE: 1997-10-22
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 3141
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (248)..(2128)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (3)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (12)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1568)..(1872)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (3126)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (3134)
;   OTHER INFORMATION: Unidentified at time of filing
US-08-956-242-1


  Query Match             2.1%;  Score 49;  DB 2;  Length 3141;
  Best Local Similarity   28.1%;  Pred. No. 0.0012;
  Matches  207;  Conservative    0;  Mismatches  520;  Indels    9;  Gaps    2;

Qy    973 tgcggcgccagtacctctatagcttttacttctccacgctgatactgactacagtggg-- 1030
          ||| |  | ||||  ||     | | ||||||  ||      |  || || |  |||||
Db   1242 tgcaggacaagtatgtcacagccctctacttcaccttcagcagcctcaccagcgtgggct 1301

Qy   1031 -cgatacaccgccgccagccagggaagaagagtacctcttcatggtgggcgacttcctgc 1089
           ||  |      ||||   ||   |    ||| |  |||||    |  ||| | | ||
Db   1302 tcggcaatgtctcgcccaacaccaactccgagaaggtcttctccatctgcgtcatgctca 1361

Qy   1090 tggccgtcatgggtttcgccaccatcatgggtagcatgagctctgtcatctacaacatga 1149
          | | |  | ||   | ||||| |||| | || | | ||  | |  ||||| |   | ||
Db   1362 tcggctccctgatgtacgccagcatcttcgggaacgtgtccgcgatcatccagcgcctgt 1421

Qy   1150 acactgcagatgcggctttctacccagatcatgcactggtgaagaagtacatgaagctgc 1209
          || | | |  |||    | | ||    |       |  || ||  ||| |||     | |
Db   1422 actcgggaactgccaggtaccacatgcagatgctgcgagtaaaagagttcattcgctttc 1481

Qy   1210 agcacgtcaaccgcaagctggagcggcgagttattgactggtatcagcacctgcagatca 1269
          | ||  ||  |  |   |||  ||  ||  ||   || | |  ||||||     ||
Db   1482 accaaatccccaaccctctgaggcaacgtcttgaagaatatttccagcacgcatggactt 1541

Qy   1270 acaagaagatgaccaacgaggtagccatcttacagcacttgcctgagcggctgcgggcag 1329
          |||  ||    |   ||  |      | | |
Db   1542 acaccaatggcattgacatg------aacatgnnnnnnnnnnnnnnnnnnnnnnnnnnnn 1595

Qy   1330 aagtggctgtgtctgtgcacctgtccactctgagccgggtgcagatctttcagaactgtg 1389

Db   1596 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 1655

Qy   1390 aggccagcctgctggaggagctggtgctgaagctgcagccccagacctactcaccaggtg 1449

Db   1656 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 1715

Qy   1450 aatatgtatgccgcaaaggagacattggccaagagatgtacatcatccgagagggtcaac 1509

Db   1716 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 1775

Qy   1510 tggccgtggtggcagatgatggtatcacacagtatgctgtgctcggtgcagggctctact 1569

Db   1776 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 1835

Qy   1570 ttggggagatcagcatcatcaacatcaaagggaacatgtctgggaaccgccgcacagcca 1629
                                                  |  ||  |    ||  ||
Db   1836 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnncacgaggggcaagtccagtgcag 1895

Qy   1630 acatcaagagcctaggttattcagacctattctgcctgagcaaggaggacctgcgggagg 1689
          || |   |   ||    || |  |||||   |   |      ||  || |||| |||||
Db   1896 acgtgcgggctctgacctactgcgacctgcacaagatccagcgggcagatctgctggagg 1955

Qy   1690 tgctgagcgagtatcc 1705
          |||||  |  ||| ||
Db   1956 tgctggacatgtaccc 1971


RESULT   9
US-09-351-215-1
; Sequence 1, Application US/09351215
; Patent No. 6087488
; GENERAL INFORMATION:
;  APPLICANT: Ganetzky, Barry S.
;  APPLICANT: Titus, Steven A.
;  TITLE OF INVENTION: Polynucleotides Encoding Herg-3
;  FILE REFERENCE: 960296.94550
;  CURRENT APPLICATION NUMBER: US/09/351,215
;  CURRENT FILING DATE: 1999-07-12
;  EARLIER APPLICATION NUMBER: 08/956,242
;  EARLIER FILING DATE: 1997-10-22
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 3141
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
```

```
;   NAME/KEY: CDS
;   LOCATION: (248)..(2128)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (3)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (12)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1568)..(1872)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (3126)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (3134)
;   OTHER INFORMATION: Unidentified at time of filing
US-09-351-215-1


  Query Match             2.1%;  Score 49;  DB 3;  Length 3141;
  Best Local Similarity   28.1%;  Pred. No. 0.0012;
  Matches  207;  Conservative    0;  Mismatches  520;  Indels    9;  Gaps    2;

Qy     973 tgcggcgccagtacctctatagcttttacttctccacgctgatactgactacagtggg-- 1030
           ||| |  | ||||  ||     | | |||||| ||      |  || || |  |||||
Db    1242 tgcaggacaagtatgtcacagccctctacttcaccttcagcagcctcaccagcgtgggct 1301

Qy    1031 -cgatacaccgccgccagccagggaagaagagtacctcttcatggtgggcgacttcctgc 1089
            ||  |      ||||   ||   |    ||| |  |||||    |  ||| | | ||
Db    1302 tcggcaatgtctcgcccaacaccaactccgagaaggtcttctccatctgcgtcatgctca 1361

Qy    1090 tggccgtcatgggtttcgccaccatcatgggtagcatgagctctgtcatctacaacatga 1149
           | | |  | ||   | ||||| |||| | || | | ||  | |  ||||| |   | ||
Db    1362 tcggctccctgatgtacgccagcatcttcgggaacgtgtccgcgatcatccagcgcctgt 1421

Cy    1150 acactgcagatgcggctttctacccagatcatgcactggtgaagaagtacatgaagctgc 1209
           || | | |  |||    | | ||    |       |  || ||  ||| |||     | |
Db    1422 actcgggaactgccaggtaccacatgcagatgctgcgagtaaaagagttcattcgctttc 1481

Qy    1210 agcacgtcaaccgcaagctggagcggcgagttattgactggtatcagcacctgcagatca 1269
           | ||  ||  |  |   |||  ||  ||  ||   || |  |  ||||||     ||
Db    1482 accaaatccccaaccctctgaggcaacgtcttgaagaatatttccagcacgcatggactt 1541

Qy    1270 acaagaagatgaccaacgaggtagccatcttacagcacttgcctgagcggctgcgggcag 1329
           ||| ||    |   ||  |      | | |
Db    1542 acaccaatggcattgacatg------aacatgnnnnnnnnnnnnnnnnnnnnnnnnnnnn 1595

Qy    1330 aagtggctgtgtctgtgcacctgtccactctgagccgggtgcagatctttcagaactgtg 1389

Db    1596 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 1655

Qy    1390 aggccagcctgctggaggagctggtgctgaagctgcagccccagacctactcaccaggtg 1449

Db    1656 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 1715

Qy    1450 aatatgtatgccgcaaaggagacattggccaagagatgtacatcatccgagagggtcaac 1509

Db    1716 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 1775

Qy    1510 tggccgtggtggcagatgatggtatcacacagtatgctgtgctcggtgcagggctctact 1569

Db    1776 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 1835

Qy    1570 ttggggagatcagcatcatcaacatcaaagggaacatgtctgggaaccgccgcacagcca 1629
                                                    |  ||  |     ||  ||
Db    1836 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 1895

Qy    1630 acatcaagagcctaggttattcagacctattctgcctgagcaaggaggacctgcgggagg 1689
           || |   |   ||    || | |||||   |    |      ||  || |||| |||||
Db    1896 acgtgcgggctctgacctactgcgacctgcacaagatccagcgggcagatctgctggagg 1955

Qy    1690 tgctgagcgagtatcc 1705
           |||||  |  ||| ||
Db    1956 tgctggacatgtaccc 1971


RESULT  10
US-09-797-906-3/c
; Sequence 3, Application US/09797906
; Patent No. 6329188
; GENERAL INFORMATION:
;  APPLICANT:  Zianghe YAN, Karen A. KETCHUM, Valentina DIFRANCESCO, Ellen M. BEASLEY
;  TITLE OF INVENTION:  ISOLATED HUMAN PROTEASE PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
;  TITLE OF INVENTION:  USES THEREOF
;  FILE REFERENCE:  CL001151CIP
;  CURRENT APPLICATION NUMBER: US/09/797,906
;  CURRENT FILING DATE:  2001-03-05
;  NUMBER OF SEQ ID NOS:  5
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 84495
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(84495)
;   OTHER INFORMATION: n = A,T,C or G
US-09-797-906-3


  Query Match             2.0%;  Score 45.2;  DB 4;  Length 84495;
  Best Local Similarity   54.9%;  Pred. No. 0.056;
  Matches   89;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy    2102 agagttgtagtaaagcctaactgctgcaactctgtcatcctgtctgcgagatcacagaca 2161
           | ||||   |  || ||| | |  |     |  ||| |||| || | ||   |  ||| |
Db   30765 ATAGTTCGTGGCAAACCTTATTTTTCTTTATAAGTCTTCCTTTCAGAGATGACCAAGATA 30706

Qy    2162 caggagcgaattggtctgtagatgcccagctagagatataggagtttaacgcacattcag 2221
             |  || ||      |  |  |    || | | | || |  ||| | |  |   |
Db   30705 TTATTGGTAAATGAAAAACAAGTTGCATATTATATACACAGCATGTTACCACTAAGAAAA 30646

Qy    2222 cccccacttaccagtacacacacacacacacacacacacaca 2263
           | | |||  ||    ||||||||||||||||||||||||||
Db   30645 CACACACACACACACACACACACACACACACACACACACACA 30604


RESULT  11
US-08-979-608A-14
; Sequence 14, Application US/08979608A
; Patent No. 6355451
;  GENERAL INFORMATION:
;       APPLICANT: Lees, Ann M.
;                  Lees, Robert S.
;                  Law, Simon W.
;                  Arjona, Anibal A.
;       TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
;                           BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
;                           TREATING ATHEROSCLEROSIS
;       NUMBER OF SEQUENCES: 42
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Fish & Richardson P.C.
;            STREET: 225 Franklin Street
;            CITY: Boston
;            STATE: MA
```

```
;             COUNTRY: USA
;             ZIP: 02110-2804
;       COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette
;             COMPUTER: IBM Compatible
;             OPERATING SYSTEM: DOS
;             SOFTWARE: FastSEQ for Windows Version 2.0
;       CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/979,608A
;             FILING DATE: 26-No. 6355451-1997
;       PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US 60/048,547
;             FILING DATE: 03-JUN-1997
;             APPLICATION NUMBER: US 60/031,930
;             FILING DATE: 27-NOV-1996
;       ATTORNEY/AGENT INFORMATION:
;             NAME: Myers, Louis
;             REGISTRATION NUMBER: 35,965
;             REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
;       TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 617/542-5070
;             TELEFAX: 617/542-8906
;   INFORMATION FOR SEQ ID NO: 14:
;       SEQUENCE CHARACTERISTICS:
;             LENGTH: 4722 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;       FEATURE:
;             NAME/KEY:  Coding Sequence
;             LOCATION:  61...1731
;       SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-08-979-608A-14


  Query Match             1.9%;  Score 44;  DB 4;  Length 4722;
  Best Local Similarity   53.5%;  Pred. No. 0.033;
  Matches   92;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy   1713 cagaccatcatggaggagaaaggacgtgagatcctgctgaaaatgaacaagttggacgtg 1772
          | ||| | |   ||||    ||  || ||   || |  |  |  | | ||| | | |
Db    496 CGGACGAGCGATGAGGTCGGAGACCGAGACCACCGGAGGCCACAGGAAAAGAAGAAGGCC 555

Qy   1773 aatgctgaggcagctgagatcgccctgcaggaggccacagagtcccggctacgaggccta 1832
          || | |  || |   |||||| | |||| |  |   |||  |  |  |||  |   || |
Db    556 AAGGGTCTGGGAAAGGAGATCACGCTGCTGATGCAGACACTGAACACGCTGAGCACCCCA 615

Qy   1833 gaccagcagctggatgatctacagaccaagtttgctcgcctcctggctgagc 1884
          ||  || ||||||  | |||    |  |||| |||    || ||  |  ||||
Db    616 GAGGAGAAGCTGGCGGCTCTGTGCAAGAAGTATGCGGAACTGCTCGAGGAGC 667


RESULT  12
US-08-205-697A-4
; Sequence 4, Application US/08205697A
; Patent No. 6218510
;  GENERAL INFORMATION:
;    APPLICANT:  Sharpe, Arlene H.
;    APPLICANT:  Borriello, Francescopaulo
;    APPLICANT:  Freeman, Gordon J.
;    APPLICANT:  Nadler, Lee M.
;    TITLE OF INVENTION:  No. 6218510el Forms of T Cell Costimulatory Molecules
;    TITLE OF INVENTION:  and Uses Therefor
;    NUMBER OF SEQUENCES:  61
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 State Street, suite 510
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/205,697A
;      FILING DATE:  02-Mar-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mandragouras, Amy E.
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  BWI-120
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  818 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..138
US-08-205-697A-4


  Query Match             1.9%;  Score 42.8;  DB 4;  Length 818;
  Best Local Similarity   56.3%;  Pred. No. 0.032;
  Matches   80;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy   2135 gtcatcctgtctgcgagatcacagacacaggagcgaattggtctgtagatgcccagctag 2194
          ||  | |||| |  ||| |  ||      ||  |||||  |  | |    |  ||  |
Db    141 GTTCTGCTGTTTCTGAGGACGTAGTTTAGAGACTGAATTCTTTGGAAAGGACATAGGGAC 200

Qy   2195 agatataggagtttaacgcacattcagccccacttaccagtacacacacacacacacac 2254
          || |     | ||    ||||||    | | |||  ||    ||||||||||||||||||
Db    201 AGTTTGCACATTTGCTTGCACATCACACACACACACACACACACACACACACACAC 260

Qy   2255 acacacacatttgctcatagac 2276
          |||||||||    | || | ||
Db    261 ACACACACACACACACACAC 282


RESULT  13
US-08-702-525-4
; Sequence 4, Application US/08702525
; Patent No. 6294660
;  GENERAL INFORMATION:
;    APPLICANT:  Sharpe, Sharpe
;    APPLICANT:  Borriello, Francescopaolo
;    APPLICANT:  Freeman, Gordon
;    APPLICANT:  Nadler, Lee
;    TITLE OF INVENTION:  No. 6294660el Forms of T Cell Costimulatory
;    TITLE OF INVENTION:  Molecules and Uses Therefor
;    NUMBER OF SEQUENCES:  65
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  28 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/702,525
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/205,697
;      FILING DATE:  02-Mar-1994
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mandragouras, Amy E.
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  BWI-120CPUS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  818 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..138
US-08-702-525-4


  Query Match             1.9%;  Score 42.8;  DB 4;  Length 818;
  Best Local Similarity   56.3%;  Pred. No. 0.032;
  Matches   80;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy   2135 gtcatcctgtctgcgagatcacagacacaggagcgaattggtctgtagatgcccagctag 2194
          ||  | |||| |  |||  |  ||      ||  |||||  |  | |    |  ||  |
Db    141 GTTCTGCTGTTTCTGAGGACGTAGTTTAGAGACTGAATTCTTTGGAAAGGACATAGGGAC 200

Qy   2195 agatataggagtttaacgcacattcagcccccacttaccagtacacacacacacacacac 2254
          || |    | ||    ||||||    | | |||  ||    ||||||||||||||||||
Db    201 AGTTTGCACATTTGCTTGCACATCACACACACACACACACACACACACACACACACAC 260

Qy   2255 acacacacatttgctcatagac 2276
          |||||||||    | || | ||
Db    261 ACACACACACACACACACAC 282


RESULT  14
PCT-US95-02576-4
; Sequence 4, Application PC/TUS9502576
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Novel Forms of T Cell Costimulatory Molecules
;    TITLE OF INVENTION:  and Uses Therefor
;    NUMBER OF SEQUENCES:  65
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 State Street, suite 510
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/02576
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/205,697
;      FILING DATE:  02-Mar-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mandragouras, Amy E.
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  BWI-120CPPC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  818 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..138
PCT-US95-02576-4


  Query Match             1.9%;  Score 42.8;  DB 5;  Length 818;
  Best Local Similarity   56.3%;  Pred. No. 0.032;
  Matches   80;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy   2135 gtcatcctgtctgcgagatcacagacacaggagcgaattggtctgtagatgcccagctag 2194
          ||  | |||| |  |||  |  ||      ||  |||||  |  | |    |  ||  |
Db    141 GTTCTGCTGTTTCTGAGGACGTAGTTTAGAGACTGAATTCTTTGGAAAGGACATAGGGAC 200

Qy   2195 agatataggagtttaacgcacattcagcccccacttaccagtacacacacacacacacac 2254
          || |    | ||    ||||||    | | |||  ||    ||||||||||||||||||
Db    201 AGTTTGCACATTTGCTTGCACATCACACACACACACACACACACACACACACACACAC 260

Qy   2255 acacacacatttgctcatagac 2276
          |||||||||    | || | ||
Db    261 ACACACACACACACACACAC 282


RESULT  15
US-08-205-697A-10
; Sequence 10, Application US/08205697A
; Patent No. 6218510
;  GENERAL INFORMATION:
;    APPLICANT:  Sharpe, Arlene H.
;    APPLICANT:  Borriello, Francescopaulo
;    APPLICANT:  Freeman, Gordon J.
;    APPLICANT:  Nadler, Lee M.
;    TITLE OF INVENTION:  No. 6218510el Forms of T Cell Costimulatory Molecules
;    TITLE OF INVENTION:  and Uses Therefor
;    NUMBER OF SEQUENCES:  61
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 State Street, suite 510
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/205,697A
;      FILING DATE:  02-Mar-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mandragouras, Amy E.
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  BWI-120
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1570 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  249..890
US-08-205-697A-10
```

Query Match 1.9%; Score 42.8; DB 4; Length 1570;
Best Local Similarity 56.3%; Pred. No. 0.043;
Matches 80; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

```
Qy   2135 gtcatcctgtctgcgagatcacagacacaggagcgaattggtctgtagatgcccagctag 2194
          ||  | |||| |  |||  |  ||       ||  |||||  |  | |    |  ||  |
Db    893 GTTCTGCTGTTTCTGAGGACGTAGTTTAGAGACTGAATTCTTTGGAAAGGACATAGGGAC 952

Qy   2195 agatataggagtttaacgcacattcagcccccacttaccagtacacacacacacacacac 2254
          || |     | ||    ||||||    | | |||  ||    ||||||||||||||||
Db    953 AGTTTGCACATTTGCTTGCACATCACACACACACACACACACACACACACACACACAC 1012

Qy   2255 acacacacatttgctcatagac 2276
          |||||||||    | || | ||
Db   1013 ACACACACACACACACACAC 1034
```

Search completed: June 20, 2002, 14:28:13
Job time: 18616 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2002, 09:07:27 ; Search time 4682.41 Seconds
(without alignments)
6652.765 Million cell updates/sec

Title: US-09-927-267-2
Perfect score: 2308
Sequence: 1 agaggggaggaggaaaacag.....agactgtgcattccatctaa 2308

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	512.6	22.2	793	10	BI830763	BI830763 603075162
2	361.6	15.7	646	9	BB625423	BB625423 BB625423
3	246.8	10.7	938	10	BM473083	BM473083 AGENCOURT
4	246	10.7	699	12	CNS04SGS	AL305173 Tetraodon
c 5	243.6	10.6	800	12	CNS03HKC	AL244389 Tetraodon
6	241.4	10.5	627	12	AQ537904	AQ537904 RPCI-11-3
7	230.8	10.0	559	10	BF190904	BF190904 237569 MA
8	227.4	9.9	560	10	BF190899	BF190899 237563 MA
9	221.6	9.6	558	9	AW640105	AW640105 bl91d05.w
c 10	217.8	9.4	621	12	AZ503692	AZ503692 1M0343L05
11	215.6	9.3	567	10	BF937274	BF937274 fm62h03.y
12	204.4	8.9	819	12	CNS02XOL	AL218622 Tetraodon
13	192.4	8.3	882	10	BG293093	BG293093 602389921
c 14	179.6	7.8	923	12	CNS042AH	AL271250 Tetraodon
c 15	167.4	7.3	580	10	BE491188	BE491188 db43c03.x
16	150.8	6.5	502	10	W84514	W84514 zd90g10.r1
17	134.6	5.8	378	12	AQ316849	AQ316849 RPCI11-97
18	132.8	5.8	449	10	W58136	W58136 zd18b02.r1
19	127.8	5.5	499	9	AW638930	AW638930 bl76e03.w
20	126.6	5.5	803	10	BG473313	BG473313 602515428
21	125.6	5.4	1019	10	BG342539	BG342539 602374710
c 22	115.6	5.0	568	10	W26690	W26690 11e4 Human
23	105.8	4.6	454	9	AI035456	AI035456 ub47c01.r
24	101.2	4.4	563	10	BG304577	BG304577 f188b05.y
25	100.6	4.4	515	10	BI706445	BI706445 fq06b04.y
c 26	100.4	4.4	743	10	BJ151852	BJ151852 BJ151852
27	95.8	4.2	1048	12	CNS025BE	AL181859 Tetraodon
28	91.2	4.0	747	10	BI733974	BI733974 603355503
29	' 90	3.9	726	12	CNS01GHZ	AL143160 Anopheles
30	89.2	3.9	573	10	BM354832	BM354832 rr17e02.y
c 31	85.2	3.7	518	9	AV605886	AV605886 AV605886
32	83.2	3.6	376	10	T97907	T97907 ye54c11.r1
33	83.2	3.6	698	10	BI548103	BI548103 603189424
c 34	76.2	3.3	471	10	BI506458	BI506458 BB170020B
c 35	76.2	3.3	471	10	BI509079	BI509079 BB170003A
c 36	75.8	3.3	410	12	BH121161	BH121161 RPCI-24-3
c 37	74.2	3.2	747	12	CNS04E7T	AL286706 Tetraodon
38	70.8	3.1	673	9	AV380874	AV380874 AV380874
39	70.6	3.1	694	9	AU137778	AU137778 AU137778
c 40	66.4	2.9	613	10	BI506874	BI506874 BB170026A
41	62.4	2.7	353	9	AL120448	AL120448 DKFZp761I
42	62.4	2.7	376	9	AL120239	AL120239 DKFZp761H
43	60.4	2.6	881	10	BE287002	BE287002 601097229
44	58.6	2.5	533	10	BJ123339	BJ123339 BJ123339
c 45	58.6	2.5	557	10	BI134074	BI134074 UI-M-BH3-

ALIGNMENTS

RESULT 1
BI830763
LOCUS BI830763 793 bp mRNA linear EST 04-OCT-2001
DEFINITION 603075162F1 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5167238 5', mRNA sequence.
ACCESSION BI830763
VERSION BI830763.1 GI:15942313
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 793)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11415 row: k column: 15
High quality sequence stop: 723.
FEATURES Location/Qualifiers
source 1. .793
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5167238"
/clone_lib="NIH_MGC_119"
/tissue_type="medulla"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source normal medulla from anonymous male age 27. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.3 kb, insert size range

0.9-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 013. Note: this is a NIH_MGC Library."
BASE COUNT 160 a 222 c 232 g 179 t
ORIGIN

Query Match 22.2%; Score 512.6; DB 10; Length 793;
Best Local Similarity 93.4%; Pred. No. 3.1e-113;
Matches 634; Conservative 0; Mismatches 34; Indels 11; Gaps 9;

```
Qy    392 caggaagttgctgcctgtcctggacccatctggggattactactactggtggctgaacac 451
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    115 CAGGAAGTTGCTGCCTGTCCTGGACCCATCTGGGGATTACTACTACTGGTGGCTGAACAC 174

Qy    452 aatggtcttcccagtcatgtataacctcatcatcctcgtgtgcagagcctgcttccccga 511
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    175 AATGGTCTTCCCAGTCATGTATAACCTCATCATCCTCGTGTGCAGAGCCTGCTTCCCCGA 234

Qy    512 cttgcagcacggttatctggtggcctggttggtgctggactacacgagtgacctgctata 571
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    235 CTTGCAGCACGGTTATCTGGTGGCCTGGTTGGTGCTGGACTACACGAGTGACCTGCTATA 294

Qy    572 cctactagacatggtggtgcgcttccacacaggattcttggaacagggcatcctggtggt 631
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    295 CCTACTAGACATGGTGGTGCGCTTCCACACAGGATTCTTGGAACAGGGCATCCTGGTGGT 354

Qy    632 ggacaagggtaggatctcgagtcgctacgttcgcacctggagtttcttc-ttggacctgg 690
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db    355 GGACAAGGGTAGGATCTCGAGTCGCTACGTTCGCACCTGGAGTTTCTTCTTTGGACCTGG 414

Qy    691 cttccctgatgcccacagatgtggtctacg-tgcggctgggcccgcacac-acccaccct 748
          |||||||||||||||||||||||||||||| |||||||||||||| |||| || ||||||
Db    415 CTTCCCTGATGCCCACAGATGTGGTCTACGTTGCGGCTGGGCCCGTACACAACACACCCT 474

Qy    749 gaggctgaaccgctttctccgcgcgccccgcctcttcgaggccttcgaccgcacagagac 808
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    475 GAGGCTGAACCGCTTTCTCCGCGCGCCCCGCCTCTTCGAGGCCTTCGACCGCACAGAGAC 534

Qy    809 ccgcacagcttacccaaatgcctttcgcattgccaagctgatgctttacatttttgtcgt 868
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db    535 CCGCACAGCTTACCCAAATGCCTTTCGCATTGCCAAGCTGATGCTTTACATTATGTCGTC 594

Qy    869 catccattggaacagctgccta-tactttgccctatcccggtacctgggcttcgggcg-t 926
          |  | | | |||||||||| ||||||||||||||||||||||||||||||||||||| |
Db    595 ATCCATTGGAAGCAGCTGCCTATTACTTTGCCCTATCCCGGTACCTGGGCTTCGGGCGTT 654

Qy    927 gacgcatgggtgtacccggaccccg--cgcagcctggctttgagcgcctgc-ggcgccag 983
          |||||||||||||||||| ||||       | |||||||||||| |||| ||||||||
Db    655 GACGCATGGGTGTACCCGGAACCCGACGCAGCCTTGGCTTTGAGCGACTGCGGGCGCCAG 714

Qy    984 tacctctatagcttttacttctccacgctg-atactgactacagtgggcg--atacaccg 1040
          |||||||||||||||||||||||||||||| ||||||||||||||||||      ||||
Db    715 TACCTCTATAGCTTTTACTTCTCCACGCTGAATACTGACTACAGTGGGCGGATACAACCG 774

Qy   1041 ccgccagccagggaagaag 1059
            || ||  |||| ||||||
Db    775 ACGGCAAGCAGGAAAGAAG 793
```

RESULT 2
BB625423
LOCUS BB625423 646 bp mRNA linear EST 26-OCT-2001
DEFINITION BB625423 RIKEN full-length enriched, adult male epididymis Mus musculus cDNA clone 9230103C20 5', mRNA sequence.
ACCESSION BB625423
VERSION BB625423.1 GI:16463693
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 646)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.

FEATURES Location/Qualifiers
source 1. .646
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="9230103C20"
/clone_lib="RIKEN full-length enriched, adult male epididymis"
/sex="male"
/tissue_type="epididymis"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:

```
                    BamHI"
BASE COUNT      146 a    221 c    170 g    109 t
ORIGIN
```

```
  Query Match             15.7%;  Score 361.6;  DB 9;  Length 646;
  Best Local Similarity   75.4%;  Pred. No. 9.6e-77;
  Matches  493;  Conservative    0;  Mismatches  149;  Indels   12;  Gaps    3;

Qy    198 aaagacaaagaagtcacagcagaagcccaacagcagcctccttcagacagtcaggcacta 257
          |||     |  ||| | |||| | ||||| |  ||||| || | |||  || ||||||||
Db      2 AAAAGACAGAAAGACGCAGCTGGAGCCCCAGGGCAGCTTCTTCCAGGAAGCCAGGCACT- 60

Qy    258 gtgcccaactccagaagtcccctacaggcagagagggtgtggacatctcacaccccagca 317
              | |  ||| | |||  |    | || |   |||| |||  |||  | || |
Db     61 ---------TTCTCAAGCCACCTCTAATGGGCGAAGCCCTGGATATCCTACAGGCTAGGA 111

Qy    318 ccagaccacagaaccatgagccaggacaccaaagtgaagacaacagagtccagtccccca 377
          ||| || |||||| |||||||||||||| |||||||||||||||||||||||  ||||||
Db    112 CCACACTACAGAAACATGAGCCAGGACAGCAAAGTGAAGACAACAGAGTCCACGCCCCCA 171

Qy    378 gccccatccaaggccaggaagttgctgcctgtcctggacccatctggggattactactac 437
          |||||| |||| |||||||||  |||||| ||||| |||||||||||||||||  ||||||
Db    172 GCCCCAACCAAAGCCAGGAAGCGGCTGCCAGTCCTAGACCCATCTGGGGATTATTACTAC 231

Qy    438 tggtggctgaacacaatggtcttcccagtcatgtataacctcatcatcctcgtgtgcaga 497
          ||| |||||||||| | || ||||||| |||||||||| |||||||||   |  |||||
Db    232 TGGGGGCTGAACACGAGGGCCTTCCCAATCATGTATAATCTCATCATCGCTGCATGCAGG 291

Qy    498 gcctgcttccccgacttgcagcacggttatctggtggcctggttggtgctggactacacg 557
          |||||||| || |||||||| | | | || ||||||||||||| | | ||||||||||||||
Db    292 GCCTGCTTTCCTGACTTGCAACCCAGATACCTGGTGGCCTGGCTTGGGCTGGACTACACG 351

Qy    558 agtgacctgctatacctactagacatggtggtgcgcttccacacaggattcttggaacag 617
          || |||||||| ||||||||||||  | | ||| ||||||||||||||||| | |||||||
Db    352 AGGGACCTGCTGTACCTACTAGATGTTGGGGTACGCTTCCACACAGGATTCCTAGAACAG 411

Qy    618 ggca-tcctggtggtggacaagggtaggatctcgagtcgctacgttcgcacctggagttt 676
          |||| |||||||||  |||||| | | |||| |  | ||||| | |||| || ||||  |
Db    412 GGCACTCCTGGTGGGGGACAAAAGCACGATCGCCCGGCGCTATGGTCGCCCCCGGAGCCT 471

Qy    677 cttcttggacctggcttccctgatgcccacagatgtggtctacgtgcggctgggcccgca 736
          | |  |||||| |||||||||| | ||| ||||||  | |||||||| | | ||||||||
Db    472 CCTGGTGGACCCGGCTTCCCTGGTCCCCCCAGATGCAGCCTACGTGCAGTTAGGCCCGCA 531

Qy    737 cacacccaccctgaggctgaaccgctttctccgcgcgccc-cgcctcttcgaggccttcg 795
          ||| |  ||| |  |||| |||| || | ||||  |||| |||||||  | ||    ||
Db    532 CACCCGTACCGTCCGGCTAAACCCGTTCCCCCGCAGGCCCTCGCCTCTGTGCGGGTCCCG 591

Qy    796 accgcacagagacccgcacagcttacccaaatgcctttcgcattgccaagctga 849
           ||| || | |||||| || ||| ||||| ||||| | |||  | ||||  |||
Db    592 CCCGGACCGGGACCCGAACCGCTCACCCACATGCCCTCCGCTATACCAACGTGA 645
```

```
RESULT   3
BM473083
LOCUS       BM473083                938 bp    mRNA    linear   EST 05-FEB-2002
DEFINITION  AGENCOURT_6466078 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5574682
            5', mRNA sequence.
ACCESSION   BM473083
VERSION     BM473083.1  GI:18522125
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 938)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM12323  row: l  column: 11
           High quality sequence start: 7
           High quality sequence stop: 706.
FEATURES             Location/Qualifiers
     source          1. .938
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5574682"
                     /clone_lib="NIH_MGC_88"
                     /tissue_type="duodenal adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: small intestine; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
                     oligo-dT primed. Average insert size 1.767 kb. Library
                     enriched for full-length clones and constructed by Life
                     Technologies. Note: this is a NIH_MGC Library."
BASE COUNT      271 a    252 c    193 g    221 t      1 others
ORIGIN
```

```
  Query Match             10.7%;  Score 246.8;  DB 10;  Length 938;
  Best Local Similarity   97.1%;  Pred. No. 6.9e-49;
  Matches  264;  Conservative    0;  Mismatches    2;  Indels    6;  Gaps    1;

Qy   2043 cccccaggtccagagtgaccccatccccatccccaggattcccacctcctagtgaatcca 2102
          | ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     32 CGCCCTGGTCCAGAGTGACCCCATCCCCATCCCCAGGATTCCCACCTCCTAGTGAATCCA 91

Qy   2103 gagttgtagtaaagcctaactgctgcaactctgtcatcctgtctgcgagatcacagacac 2162
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     92 GAGTTGTAGTAAAGCCTAACTGCTGCAACTCTGTCATCCTGTCTGCGAGATCACAGACAC 151

Qy   2163 aggagcgaattggtctgtagatgcccagctagagatataggagtttaacgcacattcagc 2222
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    152 AGGAGCGAATTGGTCTGTAGATGCCCAGCTAGAGATATAGGAGTTTAACGCACATTCAGC 211

Qy   2223 ccccacttaccagt------acacacacacacacacacacacacacatttgctcatagac 2276
          ||||||||||||||      ||||||||||||||||||||||||||||||||||||||||
Db    212 CCCCACTTACCAGTACACACACACACACACACACACACACACACATTTGCTCATAGAC 271

Qy   2277 ctgttggccccaagactgtgcattccatctaa 2308
          |||||||||||||||||||||||||||||||
Db    272 CTGTTGGCCCCAAGACTGTGCATTCCATCTAA 303
```

```
RESULT   4
CNS04SGS
LOCUS       CNS04SGS                699 bp    DNA     linear   GSS 24-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            007004 of library H from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL305173
VERSION     AL305173.1  GI:8196132
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 699)
  AUTHORS   Roest;Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
```

```
REFERENCE   2  (bases 1 to 699)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 699)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .699
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="007004"
                     /clone_lib="H"
                     /note="Genoscope sequence ID : C0BH007BH02XE1-end :
                     PUC-Ori"
BASE COUNT      162 a    216 c    151 g    166 t      4 others
ORIGIN


  Query Match             10.7%;  Score 246;  DB 12;  Length 699;
  Best Local Similarity   59.4%;  Pred. No. 9.4e-49;
  Matches  411;  Conservative    3;  Mismatches  278;  Indels    0;  Gaps    0;

Qy     602 aggattcttggaacagggcatcctggtggtggacaagggtaggatctcgagtcgctacgt 661
           ||| ||  |:||||| || || ||||||  |||| | |  | | | | |    ||||: |
Db       8 AGGCTTTCTRGAACAAGGAATGCTGGTGAAGGACCAAGCAAAGCTGTGGGCCAGCTAMAT 67

Qy     662 tcgcacctggagtttcttcttggacctggcttccctgatgcccacagatgtggtctacgt 721
           || |||      |||    ||||| ||   ||| |   |||||| ||  ||| |||| |
Db      68 CCGAACCCACCAGTTCAAACTGGACTTGCTGTCCATCCAGCCCACTGAMCTGGCCTACAT 127

Qy     722 gcggctgggcccgcacacacccaccctgaggctgaaccgctttctccgcgcgccccgcct 781
           ||||   ||||  ||    || ||| | |||||| | ||||||   || ||| |
Db     128 CTCCACGGGCATCCACAAGCCGCAGCTCAGGTTCAACCGCCTCCTCCGCTTCCCGCGCAT 187

Qy     782 cttcgaggccttcgaccgcacagagacccgcacagcttacccaaatgcctttcgcattgc 841
           |||||  ||||||  |||||  |||||  |||||    |||||  ||   ||| || |||
Db     188 GTTCGAATTCTTCGATCGCACCGAGACTCGCACCAACTACCCCAACATCTTCCGTATTTG 247

Qy     842 caagctgatgctttacatttttgtcgtcatccattggaacagctgcctatactttgccct 901
           ||| ||| |||| |||||  | ||| |||| || ||||||  |||| |||| |  ||| |
Db     248 CAACCTGGTGCTCTACATCCTGGTCATCATTCACTGGAACGCCTGCATATATTACGCCAT 307

Qy     902 atcccggtacctgggcttcgggcgtgacgcatgggtgtacccggaccccgcgcagcctgg 961
           ||||  ||   |||| |||||   ||  | ||||| | ||| |     |    || |
Db     308 ATCCAAGTCTTTGGGGTTCGGTTCGGATACTTGGGTTTTCCCAAATATTTCCATACCGGA 367

Qy     962 ctttgagcgcctgcggcgccagtacctctatagcttttacttctccacgctgatactgac 1021
           | |     ||||   ||    ||: |||| || | |||| || || || |  || ||
Db     368 ATATTCCTCCCTGACTCGGAGCTAMGTCTACTGCCTCTACTGGTCGACTCTTACTCTCAC 427

Qy    1022 tacagtgggcgatacaccgccgccagccagggaagaagagtacctcttcatggtgggcga 1081
           ||  | || || |  ||  | || | | | || ||||||||||||||  ||||    ||
Db     428 CACCATTGGAGAGATGCCCGCCCCCGTCCGAGACGAAGAGTACCTCTTTGTGGTCTTTGA 487

Qy    1082 cttcctgctggccgtcatgggtttcgccaccatcatgggtagcatgagctctgtcatcta 1141
           || ||  | | |||  ||   | |||||||| ||  |||| | | |  ||||  | ||
Db     488 TTTTCTCGTCGGCGTGCTGATCTTCGCCACAATTGTGGGGAACGTCGGCTCCATGATTGC 547

Qy    1142 caacatgaacactgcagatgcggctttctacccagatcatgcactggtgaagaagtacat 1201
           |||||||||| |  |    || |  ||| |  |        |      | ||  | |||||
Db     548 CAACATGAACGCCACCCGAGCCGAGTTCCAGGCTCGGATCGACGCCATCAAACACTACAT 607

Qy    1202 gaagctgcagcacgtcaaccgcaagctggagcggcgagttattgactggtatcagcacct 1261
           | |  | |   | ||||||    |||||||    || || ||  | ||||   |  ||||
```

```
Db     608 GCACTTCCGAAAAGTCAACAAAGAGCTGGAAACTCGCGTCATCAAGTGGTTCGACTACCT 667

Qy    1262 gcagatcaacaagaagatgaccaacgaggtag 1293
           || |||||||||        |||||  ||
Db     668 CTGGACCAACAAGAAAGCTGTGGACGAGCAAG 699


RESULT   5
CNS03HKC/c
LOCUS       CNS03HKC                800 bp    DNA     linear   GSS 17-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            027I05 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL244389
VERSION     AL244389.1  GI:7965401
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 800)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 800)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 800)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .800
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="027I05"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG027AE03SP1-end :
                     PUC-Ori"
BASE COUNT      173 a    228 c    195 g    199 t      5 others
ORIGIN


  Query Match             10.6%;  Score 243.6;  DB 12;  Length 800;
  Best Local Similarity   62.1%;  Pred. No. 3.8e-48;
  Matches  384;  Conservative    0;  Mismatches  234;  Indels    0;  Gaps    0;

Qy    1164 gctttctacccagatcatgcactggtgaagaagtacatgaagctgcagcacgtcaaccgc 1223
           ||||| |   || ||  || |    |      |       ||    | | | |||| |
Db     711 GCTTTTTTTTCATATTTTGAATCTTTTTCTTCGGGTCCACAGGCCCGGAAGGTCAGCAAG 652

Qy    1224 aagctggagcggcgagttattgactggtatcagcacctgcagatcaacaagaagatgacc 1283
           | || |||  |||||| ||  | |||| | |  |  ||  || |||   ||
Db     651 GACCTCGAGTTGCGAGTCATCAAGTGGTTTGACTATTTGTGGAACAACGGAAAAGCCCAG 592

Qy    1284 aacgaggtagccatcttacagcacttgcctgagcggctgcgggcagaagtggctgtgtct 1343
           | |||  ||   |  |   | || | || ||   |||   ||| ||  | ||  |
Db     591 GATGAGAGAGAGGTGCTGAGGTACCTTCCGGACAAGCTCAAGGCGGAGATCGCCATCCAG 532

Qy    1344 gtgcacctgtccactctgagccgggtgcagatctttcagaactgtgaggccagcctgctg 1403
```

```
          || ||| ||   |||||||     || |  || |||    |||| |||||  ||||| ||
Db    531 GTCCACATGGATACTCTGAAGAAAGTCCGTATTTTTTCCGACTGCGAGGCGGGCCTGTTG 472

Qy   1404 gaggagctggtgctgaagctgcagccccagacctactcaccaggtgaatatgtatgccgc 1463
          ||||||||||| ||||||| ||| |||   | |   || || || ||  | |||
Db    471 ATCGAGCTGGTGCTCAAGCTGCGGCCGCAGGTGTTCAGCCCCGGGGACTACATCTGCAAG 412

Qy   1464 aaaggagacattggccaagagatgtacatcatccgagagggtcaactggccgtggtggca 1523
          || || ||||||||||  ||||||||||||||   || ||  | ||||| || || ||
Db    411 AAGGGCGACATTGGCCGTGAGATGTACATCATCAAGGACGGAAAGCTGGCGGTTGTCGCC 352

Qy   1524 gatgatggtatcacacagtatgctgtgctcggtgcagggctctactttggggagatcagc 1583
          ||||||||| |||| || | ||  ||||| ||    ||    || || || |||||||||
Db    351 GATGATGGTGTCACGCAATTTGTGGTGCTGGGAAGTGGCAGTTATTTCGGCGAGATCAGC 292

Qy   1584 atcatcaacatcaaagggaacatgtctgggaaccgccgcacagccaacatcaagagccta 1643
          ||| | |||||||||||  | || | | |||||||| || || || ||||||   ||| |
Db    291 ATCCTTAACATCAAAGGCAGCAAGGCCGGGAACCGTCGAACGGCGAACATCCGCAGCATC 232

Qy   1644 ggttattcagacctattctgcctgagcaaggaggacctgcgggaggtgctgagcgagtat 1703
          ||||| || ||||| ||||||||   |||||| ||||||  ||||    ||   ||||||
Db    231 GGTTACTCGGACCTCTTCTGCCTCTCCAAGGACGACCTGATGGAGTCTTTGTTGGAGTAT 172

Qy   1704 ccacaagcacagaccatcatggaggagaaaggacgtgagatcctgctgaaaatgaacaag 1763
          ||  | ||  |  |||| |||||||||| || ||  | |||||| ||||         |
Db    171 CCGGATGCCAAAGCCATGCTGGAGGAGAAGGGCCGGCAAATCCTGATGAAGGACGGTCTG 112

Qy   1764 ttggacgtgaatgctgag 1781
          |||||| || |  | | |
Db    111 TTGGACCTGGACCCGGCG 94
```

```
RESULT   6
AQ537904
LOCUS       AQ537904                627 bp    DNA     linear   GSS 18-MAY-1999
DEFINITION  RPCI-11-326D9.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-326D9,
            DNA sequence.
ACCESSION   AQ537904
VERSION     AQ537904.1  GI:4849594
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 627)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI-11-326D9.TV
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .627
                     /organism="Homo sapiens"
                     /db_xref="GDB:7624880"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-326D9"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      151 a    164 c    157 g    155 t
ORIGIN
```

```
  Query Match             10.5%;  Score 241.4;  DB 12;  Length 627;
  Best Local Similarity   61.9%;  Pred. No. 1.2e-47;
  Matches  383;  Conservative    0;  Mismatches  236;  Indels    0;  Gaps    0;

Qy    786 gaggccttcgaccgcacagagacccgcacagcttacccaaatgcctttcgcattgccaag 845
          ||   ||| ||||||||||||||  | ||    ||||| |||   ||  | ||||  ||
Db      8 GAATTCTTTGACCGCACAGAGACAAGGACCAACTACCCCAATATGTTCAGGATTGGTAAC 67

Qy    846 ctgatgctttacatttttgtcgtcatccattggaacagctgcctatactttgccctatcc 905
          || |  | |||||| |  || ||||||| |||||   |||| | |||||||| | |||
Db     68 TTGGTCTTGTACATTCTCATCATCATCCACTGGAATGCCTGCATCTACTTTGCCATTTCC 127

Qy    906 cggtacctgggcttcgggcgtgacgcatgggtgtacccggaccccgcgcagcctggcttt 965
            || | | || || |||   ||| | ||||| |||||  ||  | |    || |    |
Db    128 AAGTTCATTGGTTTTGGGACAGACTCCTGGGTCTACCCAAACATCTCAATCCCAGAGCAT 187

Qy    966 gagcgcctgcggcgccagtacctctatagcttttacttctccacgctgatactgactaca 1025
          | ||||||      | ||||| | || ||  | ||||  |||||  |||  || || ||
Db    188 GGGCGCCTCTCCAGGAAGTACATTTACAGTCTCTACTGGTCCACCTTGACCCTTACCACC 247

Qy   1026 gtgggcgatacaccgccgccagccagggaagaagagtacctcttcatggtgggcgacttc 1085
          | || || || || || || || |  |  || || ||||| |||||  |||| |  ||||||
Db    248 ATTGGTGAGACCCCACCCCCCGTGAAAGATGAGGAGTATCTCTTTGTGGTCGTAGACTTC 307

Qy   1086 ctgctggccgtcatgggtttcgccaccatcatgggtagcatgagctctgtcatctacaac 1145
            || |||  ||  ||  ||| ||||||||  |||| |   || ||||  | ||||  ||
Db    308 TTGGTGGGTGTTCTGATTTTTGCCACCATTGTGGGCAATGTGGGCTCCATGATCTCGAAT 367

Qy   1146 atgaacactgcagatgcggctttctacccagatcatgcactggtgaagaagtacatgaag 1205
          |||||  |  ||   || |  ||| |  |  |  ||       | ||| ||||||| ||
Db    368 ATGAATGCCTCACGGGCAGAGTTCCAGGCCAAGATTGATTCCATCAAGCAGTACATGCAG 427

Qy   1206 ctgcagcacgtcaaccgcaagctggagcggcgagttattgactggtatcagcacctgcag 1265
            | |   | |||| |    |  |||||  ||| |||||     |||| | |  |||||  |
Db    428 TTCCGCAAGGTCACCAAGGACTTGGAGACGCGGGTTATCCGGTGGTTTGACTACCTGTGG 487

Qy   1266 atcaacaagaagatgaccaacgaggtagccatcttacagcacttgcctgagcggctgcgg 1325
            |||||||||||  |    | |||   |   |  |  ||  | | || ||   ||||  |
Db    488 GCCAACAAGAAGACGGTGGATGAGAAGGAGGTGCTCAAGAGCCTCCCAGACAAGCTGAAG 547

Qy   1326 gcagaagtggctgtgtctgtgcacctgtccactctgagccgggtgcagatctttcagaac 1385
          || ||  | ||  |    |||||||||  ||| ||||    ||| |  ||||| ||| ||
Db    548 GCTGAGATCGCCATCAACGTGCACCTGGACACGCTGAAGAAGGTTCGCATCTTCCAGGAC 607

Qy   1386 tgtgaggccagcctgctgg 1404
          ||||| || || |||||||
Db    608 TGTGAAGCAAGGCTGCTGG 626
```

```
RESULT   7
BF190904
LOCUS       BF190904                559 bp    mRNA    linear   EST 02-NOV-2000
DEFINITION  237569 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BF190904
VERSION     BF190904.1  GI:11074273
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 559)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
```

```
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 65  row: J  column: 10
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .559
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 2PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from testis, ovary,
                     endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT      184 a     94 c    136 g    145 t
ORIGIN


  Query Match            10.0%;  Score 230.8;  DB 10;  Length 559;
  Best Local Similarity   63.5%;  Pred. No. 4e-45;
  Matches  352;  Conservative    0;  Mismatches  202;  Indels    0;  Gaps    0;

Qy   1251 tatcagcacctgcagatcaacaagaagatgaccaacgaggtagccatcttacagcacttg 1310
          | | |  | |||  || |||||| || |      | |||  ||   ||||  || |  |
Db      1 TTTGACTATCTGTGGACCAACAAAAAAACAGTAGATGAGAAAGAAGTCTTGAAGTATCTA 60

Qy   1311 cctgagcggctgcgggcagaagtggctgtgtctgtgcacctgtccactctgagccgggtg 1370
          |||||    ||  | |||||| | ||  |    || ||| |   |||  | |    ||||
Db     61 CCTGATAAACTAAGAGCAGAAATTGCCATTAACGTTCACTTAGACACATTAAAAAAGGTG 120

Qy   1371 cagatctttcagaactgtgaggccagcctgctggaggagctggtgctgaagctgcagccc 1430
          |  || |||    ||||||||||  | ||  ||| |||| |||| ||||  | ||||||
Db    121 CGTATTTTTGCAGACTGTGAGGCTGGTCTCTTGGTGGAGTTGGTCTTGAAATTACAGCCC 180

Qy   1431 cagacctactcaccaggtgaatatgtatgccgcaaaggagacattggccaagagatgtac 1490
          |||  ||||   || || || ||  | |||   ||||| || ||||| | ||||||||||
Db    181 CAGGTCTACAGTCCTGGAGATTACATTTGCAAGAAAGGGGATATTGGACGAGAGATGTAC 240

Qy   1491 atcatccgagagggtcaactggccgtggtggcagatgatggtatcacacagtatgctgtg 1550
          || |||  ||| ||  |||| || ||||||||||||||||| ||||| |||| ||  ||
Db    241 ATTATCAAAGAAGGCAAACTTGCTGTGGTGGCAGATGATGGAATCACCCAGTTTGTGGTA 300

Qy   1551 ctcggtgcagggctctactttggggagatcagcatcatcaacatcaaagggaacatgtct 1610
           |  |||  ||   |||||||| |||||||||||| | ||||| ||||| |  |   ||
Db    301 TTGAGTGATGGCAGCTACTTTGGTGAGATCAGCATCCTTAACATTAAAGGCAGTAAAGCT 360

Qy   1611 gggaaccgccgcacagccaacatcaagagcctaggttattcagacctattctgcctgagc 1670
          || || ||  | |||||||| || || ||  | || |||||||| || ||||| ||
Db    361 GGCAATCGAAGAACAGCCAATATTAAAAGTATTGGCTATTCAGATCTGTTCTGTCTCTCG 420

Qy   1671 aaggaggacctgcgggaggtgctgagcgagtatccacaagcacagaccatcatggaggag 1730
          || || |||||   ||| |  || | |||||| ||| | ||  |    ||  | || |||
Db    421 AAAGATGACCTCATGGAAGCTCTAACCGAGTACCCAGATGCTAAATGTATGCTAGAAGAG 480

Qy   1731 aaaggacgtgagatcctgctgaaaatgaacaagttggacgtgaatgctgaggcagctgag 1790
          |||||     | ||| || |||||         | ||||  | |||  ||     ||||
Db    481 AAAGGGAAGCAAATCTTGATGAAAGATGGTCTGCTGGATATCAATATTGCAAATGCTGGA 540

Qy   1791 atcgccctgcagga 1804
          |  | ||   | ||
Db    541 AGTGACCCCAAAGA 554
```

```
RESULT   8
BF190899
LOCUS       BF190899                560 bp    mRNA    linear   EST 02-NOV-2000
DEFINITION  237563 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BF190899
VERSION     BF190899.1  GI:11074268
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 560)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 65  row: I  column: 10
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .560
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 2PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from testis, ovary,
                     endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT      183 a     95 c    136 g    146 t
ORIGIN


  Query Match             9.9%;  Score 227.4;  DB 10;  Length 560;
  Best Local Similarity   63.8%;  Pred. No. 2.7e-44;
  Matches  345;  Conservative    0;  Mismatches  196;  Indels    0;  Gaps    0;

Qy   1251 tatcagcacctgcagatcaacaagaagatgaccaacgaggtagccatcttacagcacttg 1310
          | | |  | |||  || |||||| || |      | |||  ||   ||||  || |  |
Db      1 TTTGACTATCTGTGGACCAACAAAAAAACAGTAGATGAGAAAGAAGTCTTGAAGTATCTA 60

Qy   1311 cctgagcggctgcgggcagaagtggctgtgtctgtgcacctgtccactctgagccgggtg 1370
          |||||    ||  | |||||| | ||  |    || ||| |   |||  | |    ||||
Db     61 CCTGATAAACTAAGAGCAGAAATTGCCATTAACGTTCACTTAGACACATTAAAAAAGGTG 120

Qy   1371 cagatctttcagaactgtgaggccagcctgctggaggagctggtgctgaagctgcagccc 1430
          |  || |||    ||||||||||  | ||  ||| |||| |||| ||||  | ||||||
Db    121 CGTATTTTTGCAGACTGTGAGGCTGGTCTCTTGGTGGAGTTGGTCTTGAAATTACAGCCC 180

Qy   1431 cagacctactcaccaggtgaatatgtatgccgcaaaggagacattggccaagagatgtac 1490
          |||  ||||   || || || ||  | |||   ||||| || ||||| | ||||||||||
Db    181 CAGGTCTACAGTCCTGGAGATTACATTTGCAAGAAAGGGGATATTGGACGAGAGATGTAC 240

Qy   1491 atcatccgagagggtcaactggccgtggtggcagatgatggtatcacacagtatgctgtg 1550
          || |||  ||| ||  |||| || ||||||||||||||||| ||||| |||| ||  ||
Db    241 ATTATCAAAGAAGGCAAACTTGCTGTGGTGGCAGATGATGGAATCACCCACTTTGTGGTA 300

Qy   1551 ctcggtgcagggctctactttggggagatcagcatcatcaacatcaaagggaacatgtct 1610
```

```
Db    301 TTGAGTGATGGCAGCTACTTTGGTGAGATCAGCATCCTTAACATTAAAGGCAGTAAAGCT 360

Qy   1611 gggaaccgccgcacagccaacatcaagagcctaggttattcagacctattctgcctgagc 1670
Db    361 GGCTATCGAAGAACAGCCAATATTAAAAGTATTGGCTATTCAGATCTGTTCTGTCTCTCG 420

Qy   1671 aaggaggacctgcgggaggtgctgagcgagtatccacaagcacagaccatcatggaggag 1730
Db    421 AAAGATGACCTCATGGAAGCTCTAACCGAGTACCCAGATGCTAAATGTATGCTAGAAGAG 480

Qy   1731 aaaggacgtgagatcctgctgaaaatgaacaagttggacgtgaatgctgaggcagctgag 1790
Db    481 AAAGGGAAGCAAATCTTGATGAAAGATGGTCTGCTGGATATCAATATTGCAAATGCTGGG 540

Qy   1791 a 1791
Db    541 A 541


RESULT  9
AW640105
LOCUS       AW640105                558 bp    mRNA    linear   EST 26-APR-2001
DEFINITION  bl91d05.w1 Blackshear/Soares normalized Xenopus egg library Xenopus
            laevis cDNA clone PBX0091D05 5', mRNA sequence.
ACCESSION   AW640105
VERSION     AW640105.1  GI:7397317
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 558)
  AUTHORS   Blackshear,P.J., Lai,W.S., Thorn,J.M., Kennington,E.A., Staffa,N.G.
            Jr., Moore,D.T., Bouffard,G.G., Beckstrom-Sternberg,S.M., Touchman
            ,J.W., Bonaldo,M.F. and Soares,M.B.
  TITLE     The NIEHS Xenopus maternal EST project: interim analysis of the
            first 13,879 ESTs from unfertilized eggs
  JOURNAL   Gene 267 (1), 71-87 (2001)
  MEDLINE   21211403
COMMENT     Contact: Perry J. Blackshear
            Office of Clinical Research and Laboratory of Signal Transduction
            National Institute of Environmental Health Sciences
            A2-05 NIEHS, 101 Alexander Drive, Research Triangle Park, NC 27709,
            USA
            Tel: 919 541-4899
            Fax: 919 541-4571
            Email: black009@niehs.nih.gov
            Clone is available through Research Genetics, Inc., 2130 Memorial
            Parkway, Huntsville, AL 35901
             phone 800-533-4363 ext.cdna, fax 256-536-9016 att:cdna, email
            cdna@resgen.com
             DNA Sequencing and analyses performed by National Institutes of
            Health Intramural Sequencing Center (NISC).
            PCR PRimers
            FORWARD: TGTAAAACGACGGCCAGT
            BACKWARD: CAGGAAACAGCTATGACC
            Plate: 0091  row: D  column: 05
            Seq primer: T7 primer.
FEATURES             Location/Qualifiers
     source          1. .558
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="PBX0091D05"
                     /clone_lib="Blackshear/Soares normalized Xenopus egg
                     library"
                     /sex="female"
                     /tissue_type="unfertilized egg"
                     /cell_type="unfertilized egg"
                     /dev_stage="unfertilized egg"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3-Pac; Site_1: EcoRI; Site_2: NotI;
                     PolyA-selected mRNA was prepared from unfertilized Xenopus
                     laevis eggs.  The library was constructed in the vector
                     pT7T3-Pac as described in Bonaldo, M.F., Lennon, G. and
                     Soares, M.B. 'Normalization and subtraction:  two
                     approaches to facilitate gene discovery', Genome Research
                     6:791-806, 1996.  The first strand synthesis used a
                     NotI-dT18 primer; double stranded cDNAs were ligated to
                     EcoRI adapters, digested with NotI, and directionally
                     cloned into the NotI and EcoRI-digested pT7T3-Pac vector.
                     The library contained approximately 7.2 X 10^5
                     recombinants, with average insert sizes of 1-1.5 kb."
BASE COUNT      191 a     97 c    136 g    133 t      1 others
ORIGIN


  Query Match             9.6%;  Score 221.6;  DB 9;  Length 558;
  Best Local Similarity   64.6%;  Pred. No. 6.7e-43;
  Matches  329;  Conservative    0;  Mismatches  180;  Indels    0;  Gaps    0;

Qy   1246 actggtatcagcacctgcagatcaacaagaagatgaccaacgaggtagccatcttacagc 1305
Db      6 ACGAGGTTGACTACCTCTGGACAAACAAAAAAACCATGGATGAGAAGGAGGTATTAAAGA 65

Qy   1306 acttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctgtccactctgagcc 1365
Db     66 GCTTGCCAGACAAATTAAAGGCTGAAATTGCAATTAATGTACATTTAGACACTTTAAAAA 125

Qy   1366 gggtgcagatctttcagaactgtgaggccagcctgctggaggagctggtgctgaagctgc 1425
Db    126 AAGTTAGGATCTTTCAAGACTGCGAGGCTGGATTGCTTATAGAACTGGTACTGAAACTAA 185

Qy   1426 agccccagacctactcaccaggtgaatatgtatgccgcaaaggagacattggccaagaga 1485
Db    186 GGCCCCAAGTCTTTAGCCCTGGTGACTACATTTGCCGAAAAGGAGATATTGGAAGGGAAA 245

Qy   1486 tgtacatcatccgagagggtcaactggccgtggtggcagatgatggtatcacacagtatg 1545
Db    246 TGTACATCATCAAAGAGGGGAAACTAGCTGTAGTAGCTGATGATGGCATCACTCAGTTTG 305

Qy   1546 ctgtgctcggtgcagggctctactttggggagatcagcatcatcaacatcaaagggaaca 1605
Db    306 TGGTCCTCAGTGATGGCAGTTATTTTGGCGAAATTAGCATCTTAAACATCAAAGGAAGCA 365

Qy   1606 tgtctgggaaccgccgcacagccaacatcaagagcctaggttattcagacctattctgcc 1665
Db    366 AGTCAGGCAACCGGAGGACAGCCAACATCAGAAGTATTGGCTATTCCGACCTGTTTTGTT 425

Qy   1666 tgagcaaggaggacctgcgggaggtgctgagcgagtatccacaagcacagaccatcatgg 1725
Db    426 TATCAAAGGATGATTTGATGGAGGCACTTACTGAATATCCTGATGCTAAAACAGTCTTAG 485

Qy   1726 aggagaaaggacgtgagatcctgctgaaa 1754
Db    486 AAGAAAAAGGGCGACAGATCCNCATGAAA 514


RESULT  10
AZ503692/c
LOCUS       AZ503692                621 bp    DNA     linear   GSS 05-OCT-2000
DEFINITION  1M0343L05R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0343L05 R, DNA sequence.
ACCESSION   AZ503692
VERSION     AZ503692.1  GI:10685008
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 621)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
```

```
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0343  row: L  column: 05
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 621.
FEATURES             Location/Qualifiers
     source          1. .621
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0343L05"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      143 a    166 c    123 g    189 t
ORIGIN


  Query Match             9.4%;  Score 217.8;  DB 12;  Length 621;
  Best Local Similarity   62.8%;  Pred. No. 5.9e-42;
  Matches  355;  Conservative    0;  Mismatches  207;  Indels    3;  Gaps    1;

Qy   1329 gaagtggctgtgtctgtgcacctgtccactctgagccgggtgcagatctttcagaactgt 1388
           ||| | ||  |    || || ||   |||  | |    ||| |  ||||||    |||||
Db    621 GAAATTGCCATCAACGTTCATCTAGACACGTTAAAAAAGGTTCGCATCTTTGCTGACTGT 562

Qy   1389 gaggccagcctgctggaggagctggtgctgaagctgcagccccagacctactcaccaggt 1448
           || ||  | ||| ||| |||| ||||| |||| || || ||||||   ||    || ||
Db    561 GAAGCTGGTCTGTTGGTGGAGTTGGTGTTGAAACTACAACCCCAGGTGTATAGTCCCGGA 502

Qy   1449 gaatatgtatgccgcaaaggagacattggccaagagatgtacatcatccgagagggtcaa 1508
           || ||  |||||   ||||| ||||| || |  ||||||||||||||   || ||  ||
Db    501 GATTACATATGCAAGAAAGGGGACATCGGGCGGGAGATGTACATCATCAAGGAAGGCAAA 442

Qy   1509 ctggccgtggtggcagatgatggtatcacacagtatgctgtgctcggtgcagggctctac 1568
           || || ||||||||||||||| || || ||||||| ||  ||| |  |||  ||    |||
Db    441 CTTGCTGTGGTGGCAGATGACGGAATTACACAGTTTGTGGTGTTGAGTGACGGCAGTTAC 382

Qy   1569 tttggggagatcagcatcatcaacatcaaagggaacatgtctgggaaccgccgcacagcc 1628
           ||||| ||||||||||  | || |||||||| | || | |||| |||||  | ||||||
Db    381 TTTGGTGAGATCAGCATTCTTAATATCAAAGGCAGCAAGGCTGGCAACCGAAGAACAGCC 322

Qy   1629 aacatcaagagcctaggttattcagacctattctgcctgagcaaggaggacctgcgggag 1688
           |||||| |||||| | || || || ||||| |||||||||   ||| || |||||   |||
Db    321 AACATTAAGAGCATCGGCTACTCGGACCTGTTCTGCCTCTCCAAAGATGACCTCATGGAA 262

Qy   1689 gtgctgagcgagtatccacaagcacagaccatcatggaggagaaaggacgtgagatcctg 1748
           |  || |  ||||| ||| | ||  | || ||  |||| ||||||||  |  ||||| |
Db    261 GCTCTTACAGAGTACCCAGATGCCAAAACTATGCTGGAAGAGAAAGGGAGGCAGATCTTA 202

Qy   1749 ctgaaaatgaacaagttggacgtgaatgctgaggcagctgagatcgccctgcaggaggcc 1808
            |||||        | ||||  | ||   ||        |  |  | ||   | ||
Db    201 ATGAAAGATGGTCTGCTGGATATAAACATTGCAAATATGGGCAGTGACCCTAAAGACCTG 142

Qy   1809 acagagtcccggctacgaggcctagaccagcagctggatgatctacagaccaagtttgct 1868
            ||||     | |      |  | ||   |    ||||    || || ||    |||||
Db    141 GAAGAG---AAGGTCACTCGAATGGAGGGGTCAGTGGACCTCCTGCAAACGCGATTTGCC 85

Qy   1869 cgcctcctggctgagctggagtcca 1893
           ||  || |||||||    ||||| |
Db     84 CGTATCTTGGCTGAATATGAGTCGA 60


RESULT  11
BF937274
LOCUS       BF937274                 567 bp    mRNA    linear   EST 13-FEB-2002
DEFINITION  fm62h03.y1 Zebrafish adult retina cDNA Danio rerio cDNA clone
            4200053 5' similar to SW:CNG1_BOVIN Q00194 CGMP-GATED CATION
            CHANNEL PROTEIN ;, mRNA sequence.
ACCESSION   BF937274
VERSION     BF937274.1  GI:12354594
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 567)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            Library constructed by: Chandra Tucker and Gregory Niemi DNA
            Sequencing by: Washington University Genome Sequencing Center Clone
            distribution: RessourcenZentrumPrimarDatenbank, Berlin, Germany
            (web address: www.rzpd.de)
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 420.
FEATURES             Location/Qualifiers
     source          1. .567
                     /organism="Danio rerio"
                     /strain="wild-type"
                     /db_xref="taxon:7955"
                     /clone="4200053"
                     /clone_lib="Zebrafish adult retina cDNA"
                     /sex="mixed"
                     /dev_stage="1-2 years"
                     /lab_host="E.Coli XL1-Blue MRF' (XL1-Blue MRF')"
                     /note="Vector: Lambda ZAP II (pBluescript SK-); Site_1:
                     EcoRI; Site_2: SalI; This Zebrafish library was
                     constructed by Dr. Susan E. Brockerhoff (email:
                     sbrocker@u.washington.edu) RZPD library number: 760"
BASE COUNT      156 a    133 c    143 g    134 t      1 others
ORIGIN
```

```
  Query Match             9.3%;  Score 215.6;  DB 10;  Length 567;
  Best Local Similarity   61.2%;  Pred. No. 1.9e-41;
  Matches  347;  Conservative    0;  Mismatches  220;  Indels    0;  Gaps    0;

Qy    939 tacccggaccccgcgcagcctggctttgagcgcctgcggcgccagtacctctatagcttt 998
          ||||||||||| |  | |||| |||| ||  ||   ||  | ||   ||| ||| |
Db      1 TACCCGGACCCTTCAGATCCTGAGTTTGGCCGATTGGTACGTAAATATGCCTACAGCATG 60

Qy    999 tacttctccacgctgatactgactacagtgggcgatacaccgccgccagccagggaagaa 1058
          ||||  ||||| |||| || |||||  | ||||| || |||||||| | |    |
Db     61 TACTGGTCCACCCTGACGCTCACTACCATCGGCGAAACGCCGCCGCCCGTCGAAAACTCT 120

Qy   1059 gagtacctcttcatggtgggcgacttcctgctggccgtcatgggtttcgccaccatcatg 1118
          |||||| ||||| | ||   |||||| ||  | | ||| ||  |||||| || ||| |
Db    121 GAGTACTTCTTCGTAGTCACCGACTTTCTAGTAGGCGTGTTGATTTTCGCAACAATCGTC 180

Qy   1119 ggtagcatgagctctgtcatctacaacatgaacactgcagatgcggctttctacccagat 1178
          |||| | |  |||| | |||   ||    ||| | |||   || | |||| |  |
Db    181 GGTAACGTCGGCTCAATGATCACAAATGCCAACGCAGCACGAGCCGATTTCCAAGCTCGC 240

Qy   1179 catgcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggagcggcga 1238
               |  | ||  | |||||||   | | | | |||| |    |  |||||   ||
Db    241 ATCGATGCTATTAAACAATACATGAGCTTTCGGAAAGTCACCAAGGATTTGGAGAAACGG 300

Qy   1239 gttattgactggtatcagcacctgcagatcaacaagaagatgaccaacgaggtagccatc 1298
          |||||| | |||| | |  |  ||  || ||| || |||  |  | | |||  ||   |
Db    301 GTTATTAAATGGTTTGATTATTTGTGGACCAATAAAAAGGCGGTCGATGAGAGAGAAGTA 360

Qy   1299 ttacagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctgtccact 1358
            |  || |  | || ||   |||| |||| ||  | ||  |    || || |||   ||
Db    361 CTGAAGTATCTTCCCGACAAGCTGAGGGCGGAGATAGCCATAAACGTCCATCTGGATACG 420

Qy   1359 ctgagccgggtgcagatctttcagaactgtgaggccagcctgctggaggagctggtgctg 1418
          || |     |||| |||| |  | ||||||| ||| | ||  ||| ||||||||||  |
Db    421 CTAAAAAAAGTGCGTATCTNTGCGGACTGTGAAGCCGGGCTTTTGGTGGAGCTGGTGTTA 480

Qy   1419 aagctgcagccccagacctactcaccaggtgaatatgtatgccgcaaaggagacattggc 1478
          |||||||||||| |||   |||   || || || ||| |||||   || || |||||  |
Db    481 AAGCTGCAGCCGCAGGTTTACAGCCCTGGAGACTATATATGCAAAAACGGCGACATCCGG 540

Qy   1479 caagagatgtacatcatccgagagggt 1505
          |  ||||||||||||||   ||| |||
Db    541 CGCGAGATGTACATCATTAAAGACGGT 567


RESULT  12
CNS02XOL
LOCUS       CNS02XOL                819 bp    DNA     linear   GSS 15-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            179G05 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL218622
VERSION     AL218622.1  GI:7877441
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 819)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 819)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 819)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .819
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="179G05"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG179AD03SP1-end :
                     PUC-Ori"
BASE COUNT      217 a    236 c    180 g    186 t
ORIGIN


  Query Match             8.9%;  Score 204.4;  DB 12;  Length 819;
  Best Local Similarity   54.5%;  Pred. No. 1.2e-38;
  Matches  409;  Conservative    0;  Mismatches  341;  Indels    0;  Gaps    0;

Qy    581 catggtggtgcgcttccacacaggattcttggaacagggcatcctggtggtggacaaggg 640
          | |  ||   |  || | |    | | |||||| || ||| | ||||||   |||
Db     67 CCTAATGTCACCTTTGCCCCGCAGTTACTTGGAGCAAGGCCTTCTGGTGAAAGACGCCAA 126

Qy    641 taggatctcgagtcgctacgttcgcacctggagtttcttcttggacctggcttccctgat 700
          |   |       |  |||      ||       |||   |  ||| ||  |||| ||||
Db    127 GAAACTGAGAGATAAATACCGAACAACACCCCAGTTCAAGTACGACATGATTTCCATGAT 186

Qy    701 gcccacagatgtggtctacgtgcggctgggcccgcacacacccaccctgaggctgaaccg 760
          ||||| ||  || | |   |       |||   |||  ||     |  |  | |||||
Db    187 ACCCACCGACCTGCTGTTTCTCAACTACGGCTACAACAATCCGGAGTTCCGCTTCAACCG 246

Qy    761 ctttctccgcgcgccccgcctcttcgaggccttcgaccgcacagagacccgcacagctta 820
          | ||  |     |  | || || ||    || ||  | || || ||  | ||    |
Db    247 CCTTTGCAAAATCTCAAGACTTTTTGAATTTTTTGAAAGGACTGAAACTAGAACCAGCTT 306

Qy    821 cccaaatgcctttcgcattgccaagctgatgctttacatttttgtcgtcatccattggaa 880
          ||| ||    || || ||   ||| ||  |||| |||||  | ||| |||||||| |||||
Db    307 CCCCAACATGTTCCGTATCAGCAACCTAGTGCTCTACATCCTCGTCATCATCCACTGGAA 366

Qy    881 cagctgcctatactttgccctatcccggtacctgggcttcgggcgtgacgcatgggtgta 940
          |  |||||| | ||| ||||| |||      | | ||||| ||  ||||| | ||||||||
Db    367 CGCCTGCCTGTTCTTCGCCCTCTCCAAAACCATCGGCTTTGGCAGTGACACCTGGGTGTA 426

Qy    941 cccggaccccgcgcagcctggctttgagcgcctgcggcgccagtacctctatagctttta 1000
          ||  ||  |  || || |    ||  ||||||    |  ||||| ||||     | ||
Db    427 TCCCAACATCAGCCACCCGGAGCATGGCCGCCTGGCCAGGAAGTACATCTACTCGCTGTA 486

Qy   1001 cttctccacgctgatactgactacagtgggcgatacaccgccgccagccagggaagaaga 1060
          ||  |||||||||| || || ||  | || || || ||  | || | || |||    ||
Db    487 CTGGTCCACGCTGACCCTCACCACCATCGGAGAGACGCCTGCACCCGTCAAGGATATCGA 546

Qy   1061 gtacctcttcatggtgggcgacttcctgctggccgtcatgggtttcgccaccatcatgggg 1120
          |||||  ||| |  |    ||||||||     | |||  ||   || || |  ||| | ||
Db    547 ATACCTTTTCGTTATTTCGGACTTCCTCACCGGCGTGCTGATCTTTGCTAGTATCGTCGG 606

Qy   1121 tagcatgagctctgtcatctacaacatgaacactgcagatgcggctttctacccagatca 1180
          | | |  |  |  | ||||  ||||||||  || |   || |  ||  |  || |
Db    607 AAACGTCGGTGCCATGATCTCAAACATGAATGCTTCTCGGGCCGAGTTTCAGGCAAAAAT 666

Qy   1181 tgcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggagcggcgagt 1240
          |       | ||| |||||||| || | |   | |||  |    | ||||||    | ||
Db    667 CGACTCCATAAAGCAGTACATGCAGTTCCGAAAGGTCTCCAAAGAACTGGAGGCCAGGGT 726

Qy   1241 tattgactggtatcagcacctgcagatcaacaagaagatgaccaacgaggtagccatctt 1300
          ||  | ||||   |  ||||  ||   | |||||||    | | |||   |   || |
Db    727 CATCAAGTGGTTCGACTACCTGTGGACGGAGAAGAAGACCTGCGATGAGAAGGAAGTCCT 786
```

```
Qy   1301 acagcacttgcctgagcggctgcgggcaga 1330
          || || | || ||   |||    || ||
Db    787 GAAGAACCTCCCGGACAAGCTCAAAGCCGA 816


RESULT  13
BG293093
LOCUS       BG293093                882 bp    mRNA    linear   EST 21-FEB-2001
DEFINITION  602389921F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4501732 5',
            mRNA sequence.
ACCESSION   BG293093
VERSION     BG293093.1  GI:13052518
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 882)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10369  row: j  column: 05
            High quality sequence stop: 685.
FEATURES             Location/Qualifiers
     source          1. .882
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4501732"
                     /clone_lib="NIH_MGC_94"
                     /tissue_type="retina"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 3.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      261 a    178 c    203 g    240 t
ORIGIN


  Query Match             8.3%;  Score 192.4;  DB 10;  Length 882;
  Best Local Similarity   57.9%;  Pred. No. 9.8e-36;
  Matches  360;  Conservative    0;  Mismatches  261;  Indels    1;  Gaps    1;

Qy    799 gcacagagacccgcacagcttacccaaatgcctttcgcattgccaagctgatgctttaca 858
          | || || ||  | ||    ||||||||   |||| | ||| | || ||  || | ||||
Db      3 GGACGGAAACAAGGACCAACTACCCAAACATCTTTAGGATTTCGAACCTTGTGATGTACA 62

Qy    859 tttttgtcgtcatccattggaacagctgcctatactttgccctatcccggtacctgggct 918
          |  |  || ||||||| ||||||   ||  | || | | || | ||        |||  |
Db     63 TCGTCATCATCATCCACTGGAACGCTTGTGTGTATTATTCCATCTCAAAAGCTATGGATT 122

Qy    919 tcgggcgtgacgcatgggtgtacccggaccccgcgcagcctggctttgagcgcctgcggc 978
          | |||  |||| ||||||| ||||| ||        | ||||  || |  ||  |
Db    123 T-GGGAATGACACATGGGTCTACCCTGATGTTAATGATCCTGAATTCGGCCGTTTAGCTA 181

Qy    979 gccagtacctctatagcttttacttctccacgctgatactgactacagtgggcgatacac 1038
          |  | ||  |||| ||| |||| |  || ||  |||   |||| ||  | || || || |
Db    182 GAAAATATGTCTACAGCCTTTATTGGTCTACTTTGACTTTGACAACCATTGGTGAAACTC 241

Qy   1039 cgccgccagccagggaagaagagtacctcttcatggtgggcgacttcctgctggccgtca 1098
          | || || |    |||    |||||| ||||  ||||||  |||||| |  | |  ||
Db    242 CACCTCCCGTGCTGGATTCTGAGTACATCTTTGTGGTGGTAGACTTCTTAATTGGAGTTT 301

Qy   1099 tgggtttcgccaccatcatgggtagcatgagctctgtcatctacaacatgaacactgcag 1158
          |   ||||||||||||  | |||| |||  | ||  | || | ||| |||||  | ||
Db    302 TAATTTTCGCCACCATTGTCGGTAACATAGGTTCCATGATTTCCAATATGAATGCAGCCC 361

Qy   1159 atgcggctttctacccagatcatgcactggtgaagaagtacatgaagctgcagcacgtca 1218
            || |  ||  |  ||     ||      | ||  ||||||||||  | |   | || |
Db    362 GGGCAGAATTTCAATCAAGAGTTGATGCTATCAAACAGTACATGAATTTTCGAAATGTAA 421

Qy   1219 accgcaagctggagcggcgagttattgactggtatcagcacctgcagatcaacaagaaga 1278
          |    |  ||||   | | |||||| | |||| | |  |||||  || |||||| || |
Db    422 GCAAAGATATGGAAAAGAGGGTTATTAAATGGTTTGACTACCTGTGGACCAACAAAAAAA 481

Qy   1279 tgaccaacgaggtagccatcttacagcacttgcctgagcggctgcgggcagaagtggctg 1338
            | |||  ||   || |     || | |||||    ||  |||||||| | ||
Db    482 CAGTGGATGAGAGAGAAGTCCTGAGATACCTCCCTGACAAACTAAGGGCAGAAATTGCCA 541

Qy   1339 tgtctgtgcacctgtccactctgagccgggtgcagatctttcagaactgtgaggccagcc 1398
          |    || || ||   |||  | |    ||| |  ||||||    |||||||  ||  | |
Db    542 TCAACGTTCATCTAGACACGTTAAAAAAGGTTCGCATCTTTGCTGACTGTGAAGCTGGTC 601

Qy   1399 tgctggaggagctggtgctgaa 1420
          ||  |  |||| ||||| || |
Db    602 TGTGGTGGGAGTTGGTGTTGGA 623


RESULT  14
CNS042AH/c
LOCUS       CNS042AH                923 bp    DNA     linear   GSS 18-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            077A21 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL271250
VERSION     AL271250.1  GI:7993216
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 923)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 923)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 923)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .923
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="077A21"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG077AA11LP1~end : T7"
BASE COUNT      256 a    208 c    264 g    194 t      1 others
ORIGIN
```

```
  Query Match             7.8%;  Score 179.6;  DB 12;  Length 923;
  Best Local Similarity   56.9%;  Pred. No. 1.2e-32;
  Matches  354;  Conservative    0;  Mismatches  259;  Indels    9;  Gaps    1;

Qy     601 caggattcttggaacagggcatcctggtggtggacaagggtaggatctcgagtcgctacg 660
           |||| | | ||||||| ||| | ||||||   ||| ||       ||  |    ||||||
Db     668 CAGGTTACCTGGAACAAGGCCTGCTGGTGAAAGACGAGAAGCTCCTCAGGGACGGCTACG 609

Qy     661 ttcgcacctggagtttcttcttggacctggcttccctgatgcccacagatgtggtctacg 720
           ||  || ||     ||     | ||  | | |||| || | ||||| ||  |  | |||
Db     608 TTGCCAGCTTTCAGTTTCGTCTTGATGTCGTTTCCATGCTTCCCACTGACATCTTGTACT 549

Qy     721 tgcggctgggcccgcacacacccaccctgaggctgaaccgctttctccgcgcgccccgcc 780
           || |  | ||||   ||   ||     |  |  | |||    || |  |      |||
Db     548 TGTGCTTTGGCCTCGACTACCCAGAGATCCGAATCAACAAGCTTTTGAGAATCGGCCGGA 489

Qy     781 tcttcgaggccttcgaccgcacagagacccgcacagcttacccaaatgcctttcgcattg 840
           |  | |||  ||||  | | || |||||    ||    ||||| ||   ||| |||||||
Db     488 TGATGGAGTTCTTCACCAGGACCGAGACTAAAACCAACTACCCCAACATCTTCCGCATTG 429

Qy     841 ccaagctgatgctttacatttttgtcgtcatccattggaacagctgcctatactttgccc 900
           | || |||||  | |||||  | |  |||||||  ||||||   ||| | |||||  ||
Db     428 CAAACCTGATTATGTACATCCTCATTATCATCCACTGGAACGCTTGCTTTTACTTCTCCT 369

Qy     901 tatcccggtacctgggcttcgggcgtgacgcatgggtgtaccc---------ggaccccg 951
           | |||   |   | || || ||   ||||| ||||||  |||||          || || |
Db     368 TTTCCAAATCTATCGGTTTTGGCGCTGACGAATGGGTCTACCCAGCTCTGGACGATCCGG 309

Qy     952 cgcagcctggctttgagcgcctgcggcgccagtacctctatagcttttacttctccacgc 1011
           ||||||  || | |         | |  ||||    |  ||| | ||||  || || |
Db     308 ACGAGCCTGCTTTCGGGGAGGCCATGAGGAAGTATTCTTTCAGCCTGTACTGGTCTACCC 249

Qy    1012 tgatactgactacagtgggcgatacaccgccgccagccagggaagaagagtacctcttca 1071
           |||  || || ||  | || || || || || || |||  |||    || | ||||||
Db     248 TGACCCTCACCACCATTGGAGAAACCCCCCCACCGGCCCTGGACTCGGAATTCCTCTTTC 189

Qy    1072 tggtgggcgacttcctgctggccgtcatgggtttcgccaccatcatgggtagcatgagct 1131
             |||   || |||||  | | |||| ||  ||| ||||||||  | || | |||   |
Db     188 ATGTGATTGATTTCCTAGTTGGCGTCCTGATTTTTGCCACCATTGTAGGAAACATCGCCA 129

Qy    1132 ctgtcatctacaacatgaacactgcagatgcggctttctacccagatcatgcactggtga 1191
           |  | ||||  |||||||||| | ||  | ||    ||| |  |       |      |||
Db     128 CCATGATCTCAAACATGAACGCCGCCCAGGCTCAGTTCCAGGCTCGGATCGACAACATGA 69

Qy    1192 agaagtacatgaagctgcagca 1213
           |  | ||||||  || | |||||
Db      68 AACACTACATGCAGGTACAGCA 47


RESULT 15
BE491188/c
LOCUS       BE491188                580 bp    mRNA    linear   EST 31-JUL-2000
DEFINITION  db43c03.x1 Blackshear/Soares normalized Xenopus egg library Xenopus
            laevis cDNA clone IMAGE:3301252 3' similar to gb:S42457 CGMP-GATED
            CATION CHANNEL PROTEIN (HUMAN); gb:U19717 Mus musculus cGMP-gated
            cation channel mRNA, clone mIMCD-K2 (MOUSE);, mRNA sequence.
ACCESSION   BE491188
VERSION     BE491188.1  GI:9610649
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 580)
  AUTHORS   Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.
            , Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
            ,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     WashU Xenopus EST project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Sandy Clifton, Ph.D.
            WashU Xenopus EST project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library constructed by Bento Soares and M. Fatima Bonaldo
            (University of Iowa). DNA Sequencing by: Washington University
            Genome Sequencing Center
             Clone distribution: Xenopus clones from this library are available
            through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
            Seq primer: -40UP from Gibco
            High quality sequence stop: 453.
FEATURES             Location/Qualifiers
     source          1. .580
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="IMAGE:3301252"
                     /clone_lib="Blackshear/Soares normalized Xenopus egg
                     library"
                     /sex="female"
                     /tissue_type="unfertilized egg"
                     /cell_type="unfertilized egg"
                     /dev_stage="unfertilized egg"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3-Pac; Site_1: EcoRI; Site_2: NotI;
                     PolyA-selected mRNA was prepared from unfertilized Xenopus
                     laevis eggs.  The library was constructed in the vector
                     pT7T3-Pac as described in Bonaldo, M.F., Lennon, G. and
                     Soares, M.B. 'Normalization and subtraction:  two
                     approaches to facilitate gene discovery',  Genome Research
                     6:791-806, 1996.  The first strand synthesis used a
                     NotI-dT18 primer; double stranded cDNAs were ligated to
                     EcoRI adapters, digested with NotI, and directionally
                     cloned into the NotI and EcoRI-digested pT7T3-Pac vector.
                     The library contained approximately 7.2 X 10^5
                     recombinants, with average insert sizes of 1-1.5 kb."
BASE COUNT      130 a    148 c     95 g    207 t
ORIGIN


  Query Match             7.3%;  Score 167.4;  DB 10;  Length 580;
  Best Local Similarity   66.5%;  Pred. No. 8.9e-30;
  Matches  240;  Conservative    0;  Mismatches  121;  Indels    0;  Gaps    0;

Qy    1444 caggtgaatatgtatgccgcaaaggagacattggccaagagatgtacatcatccgagagg 1503
           | ||||| ||  | ||||| |||||||| |||||    || ||||||||||||  |||||
Db     567 CTGGTGACTACATTTGCCGAAAAGGAGATATTGGAAGGGAAATGTACATCATCAAAGAGG 508

Qy    1504 gtcaactggccgtggtggcagatgatggtatcacacagtatgctgtgctcggtgcagggc 1563
           |  |||| || || || || |||||||| ||||| |||| ||  || ||| |||  ||
Db     507 GGAAACTAGCTGTAGTAGCTGATGATGGCATCACTCAGTTTGTGGTCCTCAGTGATGGCA 448

Qy    1564 tctactttggggagatcagcatcatcaacatcaaagggaacatgtctgggaaccgccgca 1623
             || ||||| || || |||||| | |||||||||||| | || ||| || |||||  | |
Db     447 GTTATTTTGGCGAAATTAGCATCTTAAACATCAAAGGAAGCAAGTCAGGCAACCGGAGGA 388

Qy    1624 cagccaacatcaagagcctaggttattcagacctattctgcctgagcaaggaggacctgc 1683
           ||||||||||||  ||  | || ||||| ||||| || ||  |    ||||| ||  ||
Db     387 CAGCCAACATCAGAAGTATTGGCTATTCCGACCTGTTTTGTTTATCAAAGGATGATTTGA 328

Qy    1684 gggaggtgctgagcgagtatccacaagcacagaccatcatggaggagaaaggacgtgaga 1743
            |||||  || |  || |||||  | ||  | ||  || | || || ||||| ||  |||
Db     327 TGGAGGCACTTACTGAATATCCTGATGCTAAAACAGTCTTAGAAGAAAAAGGGCGACAGA 268

Qy    1744 tcctgctgaaaatgaacaagttggacgtgaatgctgaggcagctgagatcgccctgcagg 1803
           ||||  |||||          | || |   | || |    ||| |    | || | |||
Db     267 TCCTCATGAAAGATGGTTTAATAGATGAAGAAGCAGCCAAAGCAGGTGCAGACCCGAAGG 208

Qy    1804 a 1804
           |
Db     207 A 207
```

Search completed: June 20, 2002, 12:16:50
Job time: 11363 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2002, 14:30:23 ; Search time 7623.77 Seconds
(without alignments)
4743.199 Million cell updates/sec

Title: US-09-927-267-3
Perfect score: 1728
Sequence: 1 atgagccaggacaccaaagt.....gacccccaggtccagagtga 1728

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1726.4	99.9	2013	6	AX355967	AX355967 Sequence
2	1726.4	99.9	2551	6	AX350404	AX350404 Sequence
3	1687.4	97.7	1835	6	AX350402	AX350402 Sequence
4	1420.8	82.2	1938	10	RNU12425	U12425 Rattus norv
5	1420.8	82.2	2351	10	RNU12623	U12623 Rattus norv
6	652.6	37.8	162965	2	AC021935	AC021935 Homo sapi
7	652.6	37.8	164434	2	AC036216	AC036216 Homo sapi
8	652.6	37.8	186609	2	AC022762	AC022762 Homo sapi
9	617.4	35.7	2641	4	BTCNGCHAN	X89600 B.taurus mR
10	617.4	35.7	3998	4	BTCNG3	X76485 B.taurus CN
11	588.2	34.0	3486	9	AF065314	AF065314 Homo sapi
12	584.8	33.8	2781	10	MMU243933	AJ243933 Mus muscu
13	583.2	33.8	3166	4	BTCAMPGC	X55010 B.taurus mR
14	578.6	33.5	1995	6	AX179746	AX179746 Sequence
15	572	33.1	2669	4	OCRACNG	X59668 O.cuniculus
16	570	33.0	2098	10	RNO272429	AJ272429 Rattus no
17	570	33.0	2212	10	RNO272428	AJ272428 Rattus no
18	570	33.0	2213	10	AB002801	AB002801 Rattus no
19	530	30.7	2815	10	AF126808	AF126808 Rattus no
20	530	30.7	3027	10	RNOLFCH	X55519 Rat mRNA fo
21	515.2	29.8	153551	9	AC092675	AC092675 Homo sapi
22	511	29.6	1428	10	MMU238241	AJ238241 Mus muscu
23	507.4	29.4	2697	5	GGCONECNG	X89598 G.gallus mR
24	499.2	28.9	3718	5	AF393839	AF393839 Oncorhync
25	498.6	28.9	104037	9	HSAF002992	AF002992 Homo sapi
26	494.8	28.6	2049	5	ICTCNC	M83111 Ictalurus p
27	490	28.4	2052	10	RNU48803	U48803 Rattus norv
28	481.2	27.8	2148	10	RNU93851	U93851 Rattus norv
29	475.6	27.5	2682	4	BTCGMPCH	X51604 B.taurus RN
30	475.4	27.5	158588	2	AC091450	AC091450 Mus muscu
31	472.2	27.3	9618	10	MMU49391	U49391 Mus musculu
32	467.6	27.1	2697	4	CFACGMP	X99914 C.familiari
33	466	27.0	2297	10	MMU19717	U19717 Mus musculu
34	466	27.0	2717	4	CFU83905	U83905 Canis famil
35	464.4	26.9	2474	10	MUSCNCG	M84742 Mus musculu
36	461.2	26.7	2061	4	SSU85404	U85404 Sus scrofa
37	461.2	26.7	2857	9	S42457	S42457 CNCG=rod ph
38	459	26.6	2612	5	GGRODCNG	X89599 G.gallus mR
39	452.4	26.2	2500	9	HUMCGMP	M84741 Human cGMP-
40	423.2	24.5	200039	2	AL671880	AL671880 Mus muscu
c 41	423.2	24.5	218872	2	AC099698	AC099698 Mus muscu
42	416.8	24.1	222364	30	AC013535	Ac013535 Homo sapi
c 43	416.8	24.1	229687	2	AC107068	AC107068 Homo sapi
44	410	23.7	2913	3	DMCNGCHAN	X89601 D.melanogas
c 45	386.4	22.4	190000	2	AC007704	AC007704 Homo sapi

ALIGNMENTS

RESULT 1
AX355967
LOCUS AX355967 2013 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 2 from Patent WO0202633.
ACCESSION AX355967
VERSION AX355967.1 GI:18620561
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Raumann,B.E., Sanjanwala,M.S., Tribouley,C.M. and Walia,N.K.
TITLE Transporters and ion channels
JOURNAL Patent: WO 0202633-A 2 10-JAN-2002;
Incyte Genomics, Inc. (US)
FEATURES Location/Qualifiers
source 1. .2013
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Incyte ID No: 6703242CB1"
BASE COUNT 472 a 576 c 557 g 408 t
ORIGIN

```
Query Match             99.9%;  Score 1726.4;  DB 6;  Length 2013;
Best Local Similarity   99.9%;  Pred. No. 0;
Matches 1727;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 atgagccaggacaccaaagtgaagacaacagagtccagtcccccagccccatccaaggcc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     124 ATGAGCCAGGACACCAAAGTGAAGACAACAGAGTCCAGTCCCCCAGCCCCATCCAAGGCC 183

Qy      61 aggaagttgctgcctgtcctggacccatctggggattactactactggtggctgaacaca 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     184 AGGAAGTTGCTGCCTGTCCTGGACCCATCTGGGGATTACTACTACTGGTGGCTGAACACA 243

Qy     121 atggtcttcccagtcatgtataacctcatcatcctcgtgtgcagagcctgcttccccgac 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     244 ATGGTCTTCCCAGTCATGTATAACCTCATCATCCTCGTGTGCAGAGCCTGCTTCCCCGAC 303

Qy     181 ttgcagcacggttatctggtggcctggttggtgctggactacacgagtgacctgctatac 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     304 TTGCAGCACGGTTATCTGGTGGCCTGGTTGGTGCTGGACTACACGAGTGACCTGCTATAC 363

Qy     241 ctactagacatggtggtgcgcttccacacaggattcttggaacagggcatcctggtggtg 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     364 CTACTAGACATGGTGGTGCGCTTCCACACAGGATTCTTGGAACAGGGCATCCTGGTGGTG 423

Qy     301 gacaagggtaggatctcgagtcgctacgttcgcacctggagtttcttcttggacctggct 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     424 GACAAGGGTAGGATCTCGAGTCGCTACGTTCGCACCTGGAGTTTCTTCTTGGACCTGGCT 483

Qy     361 tccctgatgcccacagatgtggtctacgtgcggctgggcccgcacacacccaccctgagg 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     484 TCCCTGATGCCCACAGATGTGGTCTACGTGCGGCTGGGCCCGCACACACCCACCCTGAGG 543

Qy     421 ctgaaccgctttctccgcgcgccccgcctcttcgaggccttcgaccgcacagagacccgc 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     544 CTGAACCGCTTTCTCCGCGCGCCCCGCCTCTTCGAGGCCTTCGACCGCACAGAGACCCGC 603

Qy     481 acagcttacccaaatgcctttcgcattgccaagctgatgctttacatttttgtcgtcatc 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     604 ACAGCTTACCCAAATGCCTTTCGCATTGCCAAGCTGATGCTTTACATTTTTGTCGTCATC 663

Qy     541 cattggaacagctgcctatactttgccctatcccggtacctgggcttcgggcgtgacgca 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     664 CATTGGAACAGCTGCCTATACTTTGCCCTATCCCGGTACCTGGGCTTCGGGCGTGACGCA 723

Qy     601 tgggtgtacccggaccccgcgcagcctggctttgagcgcctgcggcgccagtacctctat 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     724 TGGGTGTACCCGGACCCCGCGCAGCCTGGCTTTGAGCGCCTGCGGCGCCAGTACCTCTAT 783

Qy     661 agcttttacttctccacgctgatactgactacagtgggcgatacaccgccgccagccagg 720
           ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     784 AGCTTTTACTTCCCCACGCTGATACTGACTACAGTGGGCGATACACCGCCGCCAGCCAGG 843

Qy     721 gaagaagagtacctcttcatggtgggcgacttcctgctggccgtcatgggtttcgccacc 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     844 GAAGAAGAGTACCTCTTCATGGTGGGCGACTTCCTGCTGGCCGTCATGGGTTTCGCCACC 903

Qy     781 atcatgggtagcatgagctctgtcatctacaacatgaacactgcagatgcggctttctac 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     904 ATCATGGGTAGCATGAGCTCTGTCATCTACAACATGAACACTGCAGATGCGGCTTTCTAC 963

Qy     841 ccagatcatgcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggag 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     964 CCAGATCATGCACTGGTGAAGAAGTACATGAAGCTGCAGCACGTCAACCGCAAGCTGGAG 1023

Qy     901 cggcgagttattgactggtatcagcacctgcagatcaacaagaagatgaccaacgaggta 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1024 CGGCGAGTTATTGACTGGTATCAGCACCTGCAGATCAACAAGAAGATGACCAACGAGGTA 1083

Qy     961 gccatcttacagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctg 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1084 GCCATCTTACAGCACTTGCCTGAGCGGCTGCGGGCAGAAGTGGCTGTGTCTGTGCACCTG 1143

Qy    1021 tccactctgagccgggtgcagatctttcagaactgtgaggccagcctgctggaggagctg 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1144 TCCACTCTGAGCCGGGTGCAGATCTTTCAGAACTGTGAGGCCAGCCTGCTGGAGGAGCTG 1203

Qy    1081 gtgctgaagctgcagccccagacctactcaccaggtgaatatgtatgccgcaaaggagac 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1204 GTGCTGAAGCTGCAGCCCCAGACCTACTCACCAGGTGAATATGTATGCCGCAAAGGAGAC 1263

Qy    1141 attggccaagagatgtacatcatccgagagggtcaactggccgtggtggcagatgatggt 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1264 ATTGGCCAAGAGATGTACATCATCCGAGAGGGTCAACTGGCCGTGGTGGCAGATGATGGT 1323

Qy    1201 atcacacagtatgctgtgctcggtgcagggctctactttggggagatcagcatcatcaac 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1324 ATCACACAGTATGCTGTGCTCGGTGCAGGGCTCTACTTTGGGGAGATCAGCATCATCAAC 1383

Qy    1261 atcaaagggaacatgtctgggaaccgccgcacagccaacatcaagagcctaggttattca 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1384 ATCAAAGGGAACATGTCTGGGAACCGCCGCACAGCCAACATCAAGAGCCTAGGTTATTCA 1443

Qy    1321 gacctattctgcctgagcaaggaggacctgcgggaggtgctgagcgagtatccacaagca 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1444 GACCTATTCTGCCTGAGCAAGGAGGACCTGCGGGAGGTGCTGAGCGAGTATCCACAAGCA 1503

Qy    1381 cagaccatcatggaggagaaaggacgtgagatcctgctgaaaatgaacaagttggacgtg 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1504 CAGACCATCATGGAGGAGAAAGGACGTGAGATCCTGCTGAAAATGAACAAGTTGGACGTG 1563

Qy    1441 aatgctgaggcagctgagatcgccctgcaggaggccacagagtcccggctacgaggccta 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1564 AATGCTGAGGCAGCTGAGATCGCCCTGCAGGAGGCCACAGAGTCCCGGCTACGAGGCCTA 1623

Qy    1501 gaccagcagctggatgatctacagaccaagtttgctcgcctcctggctgagctggagtcc 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1624 GACCAGCAGCTGGATGATCTACAGACCAAGTTTGCTCGCCTCCTGGCTGAGCTGGAGTCC 1683

Qy    1561 agcgcacttaagattgcttaccgcattgaacggctggagtggcagactcgagagtggcca 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1684 AGCGCACTTAAGATTGCTTACCGCATTGAACGGCTGGAGTGGCAGACTCGAGAGTGGCCA 1743

Qy    1621 atgcccgaggacctggctgaggctgatgacgagggtgagcctgaggagggaacttccaaa 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1744 ATGCCCGAGGACCTGGCTGAGGCTGATGACGAGGGTGAGCCTGAGGAGGGAACTTCCAAA 1803

Qy    1681 gatgaagagggcagggccagccaggagggacccccaggtccagagtga 1728
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db    1804 GATGAAGAGGGCAGGGCCAGCCAGGAGGGACCCCCAGGTCCAGAGTGA 1851
```

```
RESULT   2
AX350404
LOCUS       AX350404               2551 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 29 from Patent WO0181578.
ACCESSION   AX350404
VERSION     AX350404.1  GI:18616046
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Vernet,C.A., Fernandes,E.R., Gerlach,V., Shimkets,R.A.,
            Malyankar,U.M., Boldog,F.L., Zerhusen,B.D., Spytek,K.A.,
            Majumder,K., Tchernev,V.T., Padigaru,M., Patturajan,M.,
            Burgess,C.E., Gangolli,E.A., Smithson,G., Rastelli,L.,
            Macdougall,J.R., Taupier,R.J., Grosse,W.M. and Alsobrook,J.P.
  TITLE     Novel proteins and nucleic acids encoding same
  JOURNAL   Patent: WO 0181578-A 29 01-NOV-2001;
            Curagen Corporation (US)
FEATURES             Location/Qualifiers
     source          1. .2551
                     /organism="Homo sapiens"
```

```
                     /db_xref="taxon:9606"
BASE COUNT      646 a    668 c    707 g    530 t
ORIGIN

  Query Match             99.9%;  Score 1726.4;  DB 6;  Length 2551;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1727;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 atgagccaggacaccaaagtgaagacaacagagtccagtcccccagccccatccaaggcc 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     779 ATGAGCCAGGACACCAAAGTGAAGACAACAGAGTCCAGTCCCCCAGCCCCATCCAAGGCC 838

Qy      61 aggaagttgctgcctgtcctggacccatctggggattactactactggtggctgaacaca 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     839 AGGAAGTTGCTGCCTGTCCTGGACCCATCTGGGGATTACTACTACTGGTGGCTGAACACA 898

Qy     121 atggtcttcccagtcatgtataacctcatcatcctcgtgtgcagagcctgcttccccgac 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     899 ATGGTCTTCCCAGTCATGTATAACCTCATCATCCTCGTGTGCAGAGCCTGCTTCCCCGAC 958

Qy     181 ttgcagcacggttatctggtggcctggttggtgctggactacacgagtgacctgctatac 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     959 TTGCAGCACGGTTATCTGGTGGCCTGGTTGGTGCTGGACTACACGAGTGACCTGCTATAC 1018

Qy     241 ctactagacatggtggtgcgcttccacacaggattcttggaacagggcatcctggtggtg 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1019 CTACTAGACATGGTGGTGCGCTTCCACACAGGATTCTTGGAACAGGGCATCCTGGTGGTG 1078

Qy     301 gacaagggtaggatctcgagtcgctacgttcgcacctggagtttcttcttggacctggct 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1079 GACAAGGGTAGGATCTCGAGTCGCTACGTTCGCACCTGGAGTTTCTTCTTGGACCTGGCT 1138

Qy     361 tccctgatgcccacagatgtggtctacgtgcggctgggcccgcacacacccaccctgagg 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1139 TCCCTGATGCCCACAGATGTGGTCTACGTGCGGCTGGGCCCGCACACACCCACCCTGAGG 1198

Qy     421 ctgaaccgctttctccgcgcgccccgcctcttcgaggccttcgaccgcacagagacccgc 480
             |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    1199 CTGAACCGCTTTCTCCGCGCGCCCCGCCTCTTCGAGGCCTTCGACCGCGCAGAGACCCGC 1258

Qy     481 acagcttacccaaatgcctttcgcattgccaagctgatgctttacatttttgtcgtcatc 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1259 ACAGCTTACCCAAATGCCTTTCGCATTGCCAAGCTGATGCTTTACATTTTTGTCGTCATC 1318

Qy     541 cattggaacagctgcctatactttgccctatcccggtacctgggcttcgggcgtgacgca 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1319 CATTGGAACAGCTGCCTATACTTTGCCCTATCCCGGTACCTGGGCTTCGGGCGTGACGCA 1378

Qy     601 tgggtgtacccggaccccgcgcagcctggctttgagcgcctgcggcgccagtacctctat 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1379 TGGGTGTACCCGGACCCCGCGCAGCCTGGCTTTGAGCGCCTGCGGCGCCAGTACCTCTAT 1438

Qy     661 agcttttacttctccacgctgatactgactacagtgggcgatacaccgccgccagccagg 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1439 AGCTTTTACTTCTCCACGCTGATACTGACTACAGTGGGCGATACACCGCCGCCAGCCAGG 1498

Qy     721 gaagaagagtacctcttcatggtgggcgacttcctgctggccgtcatgggtttcgccacc 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1499 GAAGAAGAGTACCTCTTCATGGTGGGCGACTTCCTGCTGGCCGTCATGGGTTTCGCCACC 1558

Qy     781 atcatgggtagcatgagctctgtcatctacaacatgaacactgcagatgcggctttctac 840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1559 ATCATGGGTAGCATGAGCTCTGTCATCTACAACATGAACACTGCAGATGCGGCTTTCTAC 1618

Qy     841 ccagatcatgcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggag 900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1619 CCAGATCATGCACTGGTGAAGAAGTACATGAAGCTGCAGCACGTCAACCGCAAGCTGGAG 1678

Qy     901 cggcgagttattgactggtatcagcacctgcagatcaacaagaagatgaccaacgaggta 960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1679 CGGCGAGTTATTGACTGGTATCAGCACCTGCAGATCAACAAGAAGATGACCAACGAGGTA 1738

Qy     961 gccatcttacagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctg 1020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1739 GCCATCTTACAGCACTTGCCTGAGCGGCTGCGGGCAGAAGTGGCTGTGTCTGTGCACCTG 1798

Qy    1021 tccactctgagccgggtgcagatctttcagaactgtgaggccagcctgctggaggagctg 1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1799 TCCACTCTGAGCCGGGTGCAGATCTTTCAGAACTGTGAGGCCAGCCTGCTGGAGGAGCTG 1858

Qy    1081 gtgctgaagctgcagccccagacctactcaccaggtgaatatgtatgccgcaaaggagac 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1859 GTGCTGAAGCTGCAGCCCCAGACCTACTCACCAGGTGAATATGTATGCCGCAAAGGAGAC 1918

Qy    1141 attggccaagagatgtacatcatccgagagggtcaactggccgtggtggcagatgatggt 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1919 ATTGGCCAAGAGATGTACATCATCCGAGAGGGTCAACTGGCCGTGGTGGCAGATGATGGT 1978

Qy    1201 atcacacagtatgctgtgctcggtgcagggctctactttggggagatcagcatcatcaac 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1979 ATCACACAGTATGCTGTGCTCGGTGCAGGGCTCTACTTTGGGGAGATCAGCATCATCAAC 2038

Qy    1261 atcaaagggaacatgtctgggaaccgccgcacagccaacatcaagagcctaggttattca 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2039 ATCAAAGGGAACATGTCTGGGAACCGCCGCACAGCCAACATCAAGAGCCTAGGTTATTCA 2098

Qy    1321 gacctattctgcctgagcaaggaggacctgcgggaggtgctgagcgagtatccacaagca 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2099 GACCTATTCTGCCTGAGCAAGGAGGACCTGCGGGAGGTGCTGAGCGAGTATCCACAAGCA 2158

Qy    1381 cagaccatcatggaggagaaaggacgtgagatcctgctgaaaatgaacaagttggacgtg 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2159 CAGACCATCATGGAGGAGAAAGGACGTGAGATCCTGCTGAAAATGAACAAGTTGGACGTG 2218

Qy    1441 aatgctgaggcagctgagatcgccctgcaggaggccacagagtcccggctacgaggccta 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2219 AATGCTGAGGCAGCTGAGATCGCCCTGCAGGAGGCCACAGAGTCCCGGCTACGAGGCCTA 2278

Qy    1501 gaccagcagctggatgatctacagaccaagtttgctcgcctcctggctgagctggagtcc 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2279 GACCAGCAGCTGGATGATCTACAGACCAAGTTTGCTCGCCTCCTGGCTGAGCTGGAGTCC 2338

Qy    1561 agcgcacttaagattgcttaccgcattgaacggctggagtggcagactcgagagtggcca 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2339 AGCGCACTTAAGATTGCTTACCGCATTGAACGGCTGGAGTGGCAGACTCGAGAGTGGCCA 2398

Qy    1621 atgcccgaggacctggctgaggctgatgacgagggtgagcctgaggagggaacttccaaa 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2399 ATGCCCGAGGACCTGGCTGAGGCTGATGACGAGGGTGAGCCTGAGGAGGGAACTTCCAAA 2458

Qy    1681 gatgaagagggcagggccagccaggagggaccccccaggtccagagtga 1728
             ||||||||||||||||||||||||||||||||||||||||||||||||
Db    2459 GATGAAGAGGGCAGGGCCAGCCAGGAGGGACCCCCAGGTCCAGAGTGA 2506


RESULT   3
AX350402
LOCUS       AX350402               1835 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 27 from Patent WO0181578.
ACCESSION   AX350402
VERSION     AX350402.1  GI:18616045
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Vernet,C.A., Fernandes,E.R., Gerlach,V., Shimkets,R.A.,
            Malyankar,U.M., Boldog,F.L., Zerhusen,B.D., Spytek,K.A.,
            Majumder,K., Tchernev,V.T., Padigaru,M., Patturajan,M.,
            Burgess,C.E., Gangolli,E.A., Smithson,G., Rastelli,L.,
            Macdougall,J.R., Taupier,R.J., Grosse,W.M. and Alsobrook,J.P.
  TITLE     Novel proteins and nucleic acids encoding same
  JOURNAL   Patent: WO 0181578-A 27 01-NOV-2001;
            Curagen Corporation (US)
```

```
FEATURES             Location/Qualifiers
     source          1. .1835
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      420 a    521 c    521 g    373 t
ORIGIN


  Query Match           97.7%;  Score 1687.4;  DB 6;  Length 1835;
  Best Local Similarity   99.4%;  Pred. No. 0;
  Matches 1727;  Conservative    0;  Mismatches    1;  Indels    9;  Gaps    3;

Qy       1 atgagccaggacaccaaagtgaagacaacagagtccagtcccccagccccatccaaggcc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      54 ATGAGCCAGGACACCAAAGTGAAGACAACAGAGTCCAGTCCCCCAGCCCCATCCAAGGCC 113

Qy      61 ---aggaagttgctgcctgtcctggacccatctggggattactactactggtggctgaac 117
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     114 AGGAGGAAGTTGCTGCCTGTCCTGGACCCATCTGGGGATTACTACTACTGGTGGCTGAAC 173

Qy     118 acaatggtcttcccagtcatgtataacctcatcatcctcgtgtgcagagcctgcttcccc 177
           |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db     174 ACAATGGTCTTCCCAGTCATGTATAACCTCATCATCCTCGTGTGCAGGGCCTGCTTCCCC 233

Qy     178 gacttgcagcacggttatctggtggcctggttggtgctggactacacgagtgacctgcta 237
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     234 GACTTGCAGCACGGTTATCTGGTGGCCTGGTTGGTGCTGGACTACACGAGTGACCTGCTA 293

Qy     238 tacctactagacatggtggtgcgcttccacaca---ggattcttggaacagggcatcctg 294
           |||||||||||||||||||||||||||||||   |||||||||||||||||||||||||
Db     294 TACCTACTAGACATGGTGGTGCGCTTCCACACAGGTGGATTCTTGGAACAGGGCATCCTG 353

Qy     295 gtggtggacaagggtaggatctcgagtcgctacgttcgcacctggagtttcttcttggac 354
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     354 GTGGTGGACAAGGGTAGGATCTCGAGTCGCTACGTTCGCACCTGGAGTTTCTTCTTGGAC 413

Qy     355 ctggcttccctgatgcccacagatgtggtctacgtgcggctgggcccgcacacacccacc 414
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     414 CTGGCTTCCCTGATGCCCACAGATGTGGTCTACGTGCGGCTGGGCCCGCACACACCCACC 473

Qy     415 ctgaggctgaaccgctttctccgcgcgccccgcctcttcgaggccttcgaccgcacagag 474
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     474 CTGAGGCTGAACCGCTTTCTCCGCGCGCCCCGCCTCTTCGAGGCCTTCGACCGCACAGAG 533

Qy     475 acccgcacagcttacccaaatgcctttcgcattgccaagctgatgctttacatttttgtc 534
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     534 ACCCGCACAGCTTACCCAAATGCCTTTCGCATTGCCAAGCTGATGCTTTACATTTTTGTC 593

Qy     535 gtcatccattggaacagctgcctatactttgccctatcccggtacctgggcttcgggcgt 594
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     594 GTCATCCATTGGAACAGCTGCCTATACTTTGCCCTATCCCGGTACCTGGGCTTCGGGCGT 653

Qy     595 gacgcatgggtgtacccggaccccgcgcagcctggctttgagcgcctgcggcgccagtac 654
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     654 GACGCATGGGTGTACCCGGACCCCGCGCAGCCTGGCTTTGAGCGCCTGCGGCGCCAGTAC 713

Qy     655 ctctatagcttttacttctccacgctgatactgactacagtgggcgatacaccgccgcca 714
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     714 CTCTATAGCTTTTACTTCTCCACGCTGATACTGACTACAGTGGGCGATACACCGCCGCCA 773

Qy     715 gccagggaagaagagtacctcttcatggtgggcgacttcctgctggccgtcatgggtttc 774
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     774 GCCAGGGAAGAAGAGTACCTCTTCATGGTGGGCGACTTCCTGCTGGCCGTCATGGGTTTC 833

Qy     775 gccaccatcatgggtagcatgagctctgtcatctacaacatgaacactgcagatgcggct 834
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     834 GCCACCATCATGGGTAGCATGAGCTCTGTCATCTACAACATGAACACTGCAGATGCGGCT 893

Qy     835 ttctacccagatcatgcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaag 894
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     894 TTCTACCCAGATCATGCACTGGTGAAGAAGTACATGAAGCTGCAGCACGTCAACCGCAAG 953

Qy     895 ctggagcggcgagttattgactggtatcagcacctgcagatcaacaagaagatgaccaac 954
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     954 CTGGAGCGGCGAGTTATTGACTGGTATCAGCACCTGCAGATCAACAAGAAGATGACCAAC 1013

Qy     955 gaggtagccatcttacagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtg 1014
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1014 GAGGTAGCCATCTTACAGCACTTGCCTGAGCGGCTGCGGGCAGAAGTGGCTGTGTCTGTG 1073

Qy    1015 cacctgtccactctgagccgggtgcagatctttcagaactgtgaggccagcctgctggag 1074
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1074 CACCTGTCCACTCTGAGCCGGGTGCAGATCTTTCAGAACTGTGAGGCCAGCCTGCTGGAG 1133

Qy    1075 gagctggtgctgaagctgcagccccagacctactcaccaggtgaatatgtatgccgcaaa 1134
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1134 GAGCTGGTGCTGAAGCTGCAGCCCCAGACCTACTCACCAGGTGAATATGTATGCCGCAAA 1193

Qy    1135 ggagacattggccaagagatgtacatcatccgagagggtcaactggccgtggtggcagat 1194
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1194 GGAGACATTGGCCAAGAGATGTACATCATCCGAGAGGGTCAACTGGCCGTGGTGGCAGAT 1253

Qy    1195 gatggtatcacacagtatgctgtgctcggtgcagggctctactttggggagatcagcatc 1254
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1254 GATGGTATCACACAGTATGCTGTGCTCGGTGCAGGGCTCTACTTTGGGGAGATCAGCATC 1313

Qy    1255 atcaacatcaaa---gggaacatgtctgggaaccgccgcacagccaacatcaagagccta 1311
           ||||||||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db    1314 ATCAACATCAAAGGTGGGAACATGTCTGGGAACCGCCGCACAGCCAACATCAAGAGCCTA 1373

Qy    1312 ggttattcagacctattctgcctgagcaaggaggacctgcgggaggtgctgagcgagtat 1371
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1374 GGTTATTCAGACCTATTCTGCCTGAGCAAGGAGGACCTGCGGGAGGTGCTGAGCGAGTAT 1433

Qy    1372 ccacaagcacagaccatcatggaggagaaaggacgtgagatcctgctgaaaatgaacaag 1431
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1434 CCACAAGCACAGACCATCATGGAGGAGAAAGGACGTGAGATCCTGCTGAAAATGAACAAG 1493

Qy    1432 ttggacgtgaatgctgaggcagctgagatcgccctgcaggaggccacagagtcccggcta 1491
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1494 TTGGACGTGAATGCTGAGGCAGCTGAGATCGCCCTGCAGGAGGCCACAGAGTCCCGGCTA 1553

Qy    1492 cgaggcctagaccagcagctggatgatctacagaccaagtttgctcgcctcctggctgag 1551
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1554 CGAGGCCTAGACCAGCAGCTGGATGATCTACAGACCAAGTTTGCTCGCCTCCTGGCTGAG 1613

Qy    1552 ctggagtccagcgcacttaagattgcttaccgcattgaacggctggagtggcagactcga 1611
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1614 CTGGAGTCCAGCGCACTTAAGATTGCTTACCGCATTGAACGGCTGGAGTGGCAGACTCGA 1673

Qy    1612 gagtggccaatgcccgaggacctggctgaggctgatgacgagggtgagcctgaggaggga 1671
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1674 GAGTGGCCAATGCCCGAGGACCTGGCTGAGGCTGATGACGAGGGTGAGCCTGAGGAGGGA 1733

Qy    1672 acttccaaagatgaagagggcagggccagccaggagggaccccccaggtccagagtga 1728
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1734 ACTTCCAAAGATGAAGAGGGCAGGGCCAGCCAGGAGGGACCCCCAGGTCCAGAGTGA 1790


RESULT   4
RNU12425
LOCUS       RNU12425                1938 bp    mRNA    linear   ROD 05-APR-1995
DEFINITION  Rattus norvegicus olfactory cyclic nucleotide-gated channel mRNA,
            complete cds.
ACCESSION   U12425
VERSION     U12425.1  GI:548083
KEYWORDS    .
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1938)
  AUTHORS   Liman,E.R. and Buck,L.B.
  TITLE     A second subunit of the olfactory cyclic nucleotide-gated channel
            confers high sensitivity to cAMP
```

```
  JOURNAL   Neuron 13 (3), 611-621 (1994)
  MEDLINE   95000663
REFERENCE   2  (bases 1 to 1938)
  AUTHORS   Liman,E.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-1994) Emily R. Liman, Neurobiology, Harvard
            Medical School, 220 Longwood Avenue, Boston, MA 02115, USA
FEATURES             Location/Qualifiers
     source          1. .1938
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /tissue_type="olfactory neuroepithelium"
     mRNA            1. .1937
                     /evidence=experimental
     CDS             107. .1834
                     /codon_start=1
                     /product="olfactory cyclic nucleotide-gated channel"
                     /protein_id="AAA64748.1"
                     /db_xref="GI:548084"
                     /translation="MSQDGKVKTTESTPPAPTKARKWLPVLDPSGDYYYWWLNTMVFP
                     IMYNLIIVVCRACFPDLQHSYLVAWFVLDYTSDLLYLLDIGVRFHTGFLEQGILVVDK
                     GMIASRYVRTWSFLLDLASLVPTDAAYVQLGPHIPTLRLNRFLRVPRLFEAFDRTETR
                     TAYPNAFRIAKLMLYIFVVIHWNSCLYFALSRYLGFGRDAWVYPDPAQPGFERLRRQY
                     LYSFYFSTLILTTVGDTPLPDREEEYLFMVGDFLLAVMGFATIMGSMSSVIYNMNTAD
                     AAFYPDHALVKKYMKLQHVNKRLERRVIDWYQHLQINKKMTNEVAILQHLPERLRAEV
                     AVSVHLSTLSRVQIFQNCEASLLEELVLKLQPQTYSPGEYVCRKGDIGREMYIIREGQ
                     LAVVADDGVTQYAVLGAGLYFGEISIINIKGNMSGNRRTANIKSLGYSDLFCLSKEDL
                     REVLSEYPQAQAVMEEKGREILLKMNKLDVNAEAAEIALQEATESRLKGLDQQLDDLQ
                     TKFARLLAELESSALKIAYRIERLEWQTREWPMPEDMGEADDEAEPGEGTSKDGEKKA
                     GQAGPSGIE"
     polyA_site      1938
                     /note="3 A residues"
BASE COUNT      458 a    531 c    543 g    406 t
ORIGIN
```

Query Match 82.2%; Score 1420.8; DB 10; Length 1938;
Best Local Similarity 88.9%; Pred. No. 0;
Matches 1536; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

```
Qy       1 atgagccaggacaccaaagtgaagacaacagagtccagtcccccagccccatccaaggcc 60
           ||||||||||||  ||||||||||| ||||||||||  ||||||||||| |||| |||
Db     107 ATGAGCCAGGACGGCAAAGTGAAGACCACAGAGTCCACGCCCCCAGCCCCAACCAAAGCC 166

Qy      61 aggaagttgctgcctgtcctggacccatctggggattactactactggtggctgaacaca 120
           ||||||| |||||| ||||| ||||||||||||||||||||||||||||||||||||||||
Db     167 AGGAAGTGGCTGCCAGTCCTAGACCCATCTGGGGATTACTACTACTGGTGGCTGAACACA 226

Qy     121 atggtcttcccagtcatgtataacctcatcatcctcgtgtgcagagcctgcttccccgac 180
           |||||||||||| ||||||| |||||||||||| | || ||||| |||||||| || |||
Db     227 ATGGTCTTCCCAATCATGTACAACCTCATCATCGTTGTATGCAGGGCCTGCTTTCCTGAC 286

Qy     181 ttgcagcacggttatctggtggcctggttggtgctggactacacgagtgacctgctatac 240
           |||||||| |||| |||||||||||||| ||||||||||||||||||||||||||| |||
Db     287 TTGCAGCACAGTTACCTGGTGGCCTGGTTCGTGCTGGACTACACGAGTGACCTGCTGTAC 346

Qy     241 ctactagacatggtggtgcgcttccacacaggattcttggaacagggcatcctggtggtg 300
           |||||||| || | ||| |||||||||||||||||| | || ||||||||||||||||||
Db     347 CTACTAGATATTGGGGTACGCTTCCACACAGGATTCCTAGAGCAGGGCATCCTGGTGGTA 406

Qy     301 gacaagggtaggatctcgagtcgctacgttcgcacctggagtttcttcttggacctggct 360
           ||||| || | |||| | |||||||||| |||||||||| ||| | ||||||||||
Db     407 GACAAAGGCATGATCGCCAGTCGCTACGTCCGCACCTGGAGCTTCCTGTTGGACCTGGCG 466

Qy     361 tccctgatgcccacagatgtggtctacgtgcggctgggcccgcacacacccaccctgagg 420
           |||||| | |||||||||| || ||| |||| ||||||||| |||| || || ||  ||
Db     467 TCCCTGGTCCCCACAGATGCGGCCTATGTGCAGCTGGGCCCCCACATCCCTACACTCCGG 526

Qy     421 ctgaaccgctttctccgcgcgccccgcctcttcgaggccttcgaccgcacagagacccgc 480
           || |||||||||||||| | |||||||||||||||||| || || || ||||||||||||
Db     527 CTAAACCGCTTTCTCCGAGTGCCCCGCCTCTTCGAGGCTTTTGATCGTACAGAGACCCGC 586

Qy     481 acagcttacccaaatgcctttcgcattgccaagctgatgctttacatttttgtcgtcatc 540
           || |||||||||||||||| ||||| |||||||||||||||||||||||||||||| ||||||
Db     587 ACGGCTTACCCAAATGCCTTCCGCATAGCCAAGCTGATGCTTTACATTTTTGTTGTCATC 646

Qy     541 cattggaacagctgcctatactttgccctatcccggtacctgggcttcgggcgtgacgca 600
           |||||||||| ||| ||||||| ||||| ||| |||||||||||||| || || || ||
Db     647 CATTGGAACAGTTGCTTATACTTCGCCCTGTCCAGGTACCTGGGCTTTGGACGGGATGCG 706

Qy     601 tgggtgtacccggaccccgcgcagcctggctttgagcgcctgcggcgccagtacctctat 660
           ||||| ||||| ||||| ||||| |||||||||||||| ||||||||||||| ||||||
Db     707 TGGGTATACCCAGACCCTGCGCAACCTGGCTTTGAGCGCTTGCGGCGCCAGTATCTCTAC 766

Qy     661 agcttttacttctccacgctgatactgactacagtgggcgatacaccgccgccagccagg 720
           ||||| |||||||||| ||||| ||||| |||||||| || || |||| ||||| | |
Db     767 AGCTTCTACTTCTCCACTCTGATCCTGACCACAGTGGGTGACACGCCGCTGCCAGACCGA 826

Qy     721 gaagaagagtacctcttcatggtgggcgacttcctgctggccgtcatgggtttcgccacc 780
           || ||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db     827 GAGGAAGAGTACCTCTTCATGGTGGGTGACTTCCTGCTGGCCGTCATGGGTTTCGCCACC 886

Qy     781 atcatgggtagcatgagctctgtcatctacaacatgaacactgcagatgcggctttctac 840
           |||||||||||||||||||||||||| ||||||||||||||||||||||||| ||||||
Db     887 ATCATGGGTAGCATGAGCTCTGTCATTTACAACATGAACACTGCAGATGCGGCCTTCTAC 946

Qy     841 ccagatcatgcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggag 900
           ||||| ||||| ||||| |||||||||||||||||||||||| ||||||   | |||||||
Db     947 CCAGACCATGCGCTGGTAAAGAAGTACATGAAGCTGCAGCATGTCAACAAGAGGCTGGAG 1006

Qy     901 cggcgagttattgactggtatcagcacctgcagatcaacaagaagatgaccaacgaggta 960
           |||||||||||||||||||||| ||||| || ||||||||||||||||||||||||||||
Db    1007 CGGCGAGTTATTGACTGGTACCAGCATCTTCAGATCAACAAGAAGATGACCAACGAGGTA 1066

Qy     961 gccatcttacagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctg 1020
           |||||||| |||||| |||||||||||||||||||| || || |||||||| |||||||||
Db    1067 GCCATCTTGCAGCACCTGCCTGAGCGGCTGCGGGCGGAGGTTGCTGTGTCCGTGCACCTG 1126

Qy    1021 tccactctgagccgggtgcagatctttcagaactgtgaggccagcctgctggaggagctg 1080
           || || |||||||| || |||||||| |||||||||| |||||||||||||| ||||||
Db    1127 TCTACCCTGAGCCGAGTACAGATCTTCCAGAACTGTGAAGCCAGCCTGCTGGAAGAGCTG 1186

Qy    1081 gtgctgaagctgcagccccagacctactcaccaggtgaatatgtatgccgcaaaggagac 1140
           |||||||||| |||||||||||||||||| ||||| |||||||| |||||||||| |||
Db    1187 GTGCTGAAGCTACAGCCCCAGACCTACTCGCCAGGCGAATATGTGTGCCGCAAAGGGGAC 1246

Qy    1141 attggccaagagatgtacatcatccgagagggtcaactggccgtggtggcagatgatggt 1200
           ||||||| |||||||||||||||||| ||||| || ||||| |||||||| |||||||||
Db    1247 ATTGGCCGAGAGATGTACATCATCCGTGAGGGCCAGCTGGCTGTGGTGGCCGATGATGGT 1306

Qy    1201 atcacacagtatgctgtgctcggtgcagggctctactttggggagatcagcatcatcaac 1260
           |||||||||||||||||| |||||||||||||||||||||||||||||| |||||||||
Db    1307 GTCACACAGTATGCTGTGCTTGGTGCAGGGCTCTACTTTGGGGAGATCAGTATCATCAAC 1366

Qy    1261 atcaaagggaacatgtctgggaaccgccgcacagccaacatcaagagcctaggttattca 1320
           |||||||||||||||||||| ||||| || ||||||||||||||||||||||||||||||
Db    1367 ATCAAAGGGAACATGTCTGGAAACCGACGAACAGCCAACATCAAGAGCCTAGGTTATTCA 1426

Qy    1321 gacctattctgcctgagcaaggaggacctgcgggaggtgctgagcgagtatccacaagca 1380
           ||||| |||||||| |||||||||| |||||||||| ||||| ||||| ||||| ||
Db    1427 GACCTGTTCTGCCTCAGCAAGGAGGATCTGCGGGAGGTACTGAGTGAGTACCCACAGGCC 1486

Qy    1381 cagaccatcatggaggagaaaggacgtgagatcctgctgaaaatgaacaagttggacgtg 1440
           ||| |  |||||||||||| || || || ||| |||| |||||||| |||||||| |||
Db    1487 CAGGCGGTCATGGAGGAGAAGGGCCGAGAAATCTTGCTCAAAATGAATAAGTTGGATGTG 1546

Qy    1441 aatgctgaggcagctgagatcgccctgcaggaggccacagagtcccggctacgaggccta 1500
           |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db    1547 AATGCTGAGGCAGCTGAGATCGCCCTCCAGGAGGCCACAGAGTCTCGGCTCAAAGGCCTC 1606

Qy    1501 gaccagcagctggatgatctacagaccaagtttgctcgcctcctggctgagctggagtcc 1560
           |||||||||| |||||||| |||||||||||||||||||| ||||||||||||||||||
Db    1607 GACCAGCAGCTTGATGATCTGCAGACCAAGTTTGCTCGCCTACTGGCTGAGCTGGAGTCC 1666

Qy    1561 agcgcacttaagattgcttaccgcattgaacggctggagtggcagactcgagagtggcca 1620
           || ||||| ||||| |||||||||||| ||| ||||||||||||||||||||||||||||
```

```
Db    1667 AGTGCACTGAAGATAGCTTACCGCATCGAAAGGCTGGAGTGGCAGACTCGAGAGTGGCCA 1726

Qy    1621 atgcccgaggacctggctgaggctgatgacgagggtgagcctgaggagggaacttccaaa 1680
           ||||| |||||| ||| |||||||||||| |||| ||||||||  || || || |||||
Db    1727 ATGCCAGAGGACATGGGTGAGGCTGATGATGAGGCTGAGCCTGGAGAAGGGACGTCCAAG 1786

Qy    1681 gatgaagagggcagggccagccaggagggacccccaggtccagagtga 1728
           |||| |||||| | |||  |||||| ||||||| ||||   ||| |||
Db    1787 GATGGAGAGGGAAAGGCTGGCCAGGCGGGACCCTCAGGCATAGAATGA 1834


RESULT  5
RNU12623
LOCUS       RNU12623                2351 bp    mRNA    linear   ROD 15-SEP-1994
DEFINITION  Rattus norvegicus Sprague-Dawley cyclic nucleotide gated cation
            channel mRNA, complete cds.
ACCESSION   U12623
VERSION     U12623.1  GI:538128
KEYWORDS    .
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 2351)
  AUTHORS   Bradley,J., Li,J., Davidson,N., Lester,H.A. and Zinn,K.
  TITLE     Heteromeric olfactory cyclic nucleotide-gated channels: A new
            subunit that confers increased sensitivity to cAMP
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 91, 8890-8894 (1994)
  MEDLINE   94377458
REFERENCE   2  (bases 1 to 2351)
  AUTHORS   Bradley,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-1994) Jonathan Bradley, Biology, California
            Institute of Technology, Pasadena, CA 91125, USA
COMMENT     On Sep 15, 1994 this sequence version replaced gi:521147.
FEATURES             Location/Qualifiers
     source          1. .2351
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
     CDS             109. .1836
                     /codon_start=1
                     /product="cyclic nucleotide gated cation channel"
                     /protein_id="AAA21464.1"
                     /db_xref="GI:538129"
                     /translation="MSQDGKVKTTESTPPAPTKARKWLPVLDPSGDYYYWWLNTMVFP
                     IMYNLIIVVCRACFPDLQHSYLVAWFVLDYTSDLLYLLDIGVRFHTGFLEQGILVVDK
                     GMIASRYVRTWSFLLDLASLVPTDAAYVQLGPHIPTLRLNRFLRVPRLFEAFDRTETR
                     TAYPNAFRIAKLMLYIFVVIHWNSCLYFALSRYLGFGRDAWVYPDPAQPGFERLRRQY
                     LYSFYFSTLILTTVGDTPLPDREEEYLFMVGDFLLAVMGFATIMGSMSSVIYNMNTAD
                     AAFYPDHALVKKYMKLQHVNKRLERRVIDWYQHLQINKKMTNEVAILQHLPERLRAEV
                     AVSVHLSTLSRVQIFQNCEASLLEELVLKLQPQTYSPGEYVCRKGDIGREMYIIREGQ
                     LAVVADDGVTQYAVLGAGLYFGEISIINIKGNMSGNRRTANIKSLGYSDLFCLSKEDL
                     REVLSEYPQAQAVMEEKGREILLKMNKLDVNAEAAEIALQEATESRLKGLDQQLDDLQ
                     TKFARLLAELESSALKIAYRIERLEWQTREWPMPEDMGEADDEAEPGEGTSKDGEGKA
                     GQAGPSGIE"
BASE COUNT      580 a    624 c    637 g    510 t
ORIGIN


  Query Match           82.2%;  Score 1420.8;  DB 10;  Length 2351;
  Best Local Similarity  88.9%;  Pred. No. 0;
  Matches 1536;  Conservative    0;  Mismatches  192;  Indels    0;  Gaps    0;

Qy      1 atgagccaggacaccaaagtgaagacaacagagtccagtcccccagccccatccaaggcc 60
          |||||||||||  |||||||||| ||||||||||  |||| |||
Db    109 ATGAGCCAGGACGGCAAAGTGAAGACCACAGAGTCCACGCCCCCAGCCCCAACCAAAGCC 168

Qy     61 aggaagttgctgcctgtcctggacccatctggggattactactactggtggctgaacaca 120
          |||||||  |||||| ||||| ||||||||||||||||||||||||||||||||||||||
Db    169 AGGAAGTGGCTGCCAGTCCTAGACCCATCTGGGGATTACTACTACTGGTGGCTGAACACA 228

Qy    121 atggtcttcccagtcatgtataacctcatcatcctcgtgtgcagagcctgcttccccgac 180
          |||||||||||| ||||||| |||||||||||| | || ||||| |||||||| || |||
Db    229 ATGGTCTTCCCAATCATGTACAACCTCATCATCGTTGTATGCAGGGCCTGCTTTCCTGAC 288

Qy    181 ttgcagcacggttatctggtggcctggttggtgctggactacacgagtgacctgctatac 240
          |||||||||  |||| |||||||||||||| ||||||||||||||||||||||||||| |||
Db    289 TTGCAGCACAGTTACCTGGTGGCCTGGTTCGTGCTGGACTACACGAGTGACCTGCTGTAC 348

Qy    241 ctactagacatggtggtgcgcttccacacaggattcttggaacagggcatcctggtggtg 300
          ||||||| || | ||| |||||||||||||||||| | || ||||||||||||||||||
Db    349 CTACTAGATATTGGGGTACGCTTCCACACAGGATTCCTAGAGCAGGGCATCCTGGTGGTA 408

Qy    301 gacaagggtaggatctcgagtcgctacgttcgcacctggagtttcttcttggacctggct 360
          ||||| || | |||| | |||||||||||| |||||||||| ||| | ||||||||||
Db    409 GACAAAGGCATGATCGCCAGTCGCTACGTCCGCACCTGGAGCTTCCTGTTGGACCTGGCG 468

Qy    361 tccctgatgcccacagatgtggtctacgtgcggctgggcccgcacacacccaccctgagg 420
          ||||||  | |||||||||| || ||| |||| |||||||||  |||| || || ||  ||
Db    469 TCCCTGGTCCCCACAGATGCGGCCTATGTGCAGCTGGGCCCCCACATCCCTACACTCCGG 528

Qy    421 ctgaaccgctttctccgcgcgccccgcctcttcgaggccttcgaccgcacagagacccgc 480
          || |||||||||||||| | |||||||||||||||||| || || || ||||||||||||
Db    529 CTAAACCGCTTTCTCCGAGTGCCCCGCCTCTTCGAGGCTTTTGATCGTACAGAGACCCGC 588

Qy    481 acagcttacccaaatgcctttcgcattgccaagctgatgctttacatttttgtcgtcatc 540
          || |||||||||||||||| ||||| |||||||||||||||||||||||||||||||||
Db    589 ACGGCTTACCCAAATGCCTTCCGCATAGCCAAGCTGATGCTTTACATTTTTGTTGTCATC 648

Qy    541 cattggaacagctgcctatactttgccctatcccggtacctgggcttcgggcgtgacgca 600
          |||||||||| ||| ||||||| ||||| ||| ||||||||||||| || || || ||
Db    649 CATTGGAACAGTTGCTTATACTTCGCCCTGTCCAGGTACCTGGGCTTTGGACGGGATGCG 708

Qy    601 tgggtgtacccggaccccgcgcagcctggctttgagcgcctgcggcgccagtacctctat 660
          ||||| ||||| ||||| |||| |||||||||||||| |||||||||||| |||||||||
Db    709 TGGGTATACCCAGACCCTGCGCAACCTGGCTTTGAGCGCTTGCGGCGCCAGTATCTCTAC 768

Qy    661 agcttttacttctccacgctgatactgactacagtgggcgatacaccgccgccagccagg 720
          ||||| |||||||||| ||||| ||||| |||||||| || || |||| ||||| | |
Db    769 AGCTTCTACTTCTCCACTCTGATCCTGACCACAGTGGGTGACACGCCGCTGCCAGACCGA 828

Qy    721 gaagaagagtacctcttcatggtgggcgacttcctgctggccgtcatgggtttcgccacc 780
          || |||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db    829 GAGGAAGAGTACCTCTTCATGGTGGGTGACTTCCTGCTGGCCGTCATGGGTTTCGCCACC 888

Qy    781 atcatgggtagcatgagctctgtcatctacaacatgaacactgcagatgcggctttctac 840
          |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db    889 ATCATGGGTAGCATGAGCTCTGTCATTTACAACATGAACACTGCAGATGCGGCCTTCTAC 948

Qy    841 ccagatcatgcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggag 900
          ||||| ||||| ||||| |||||||||||||||||||||||| ||||||   | |||||||
Db    949 CCAGACCATGCGCTGGTAAAGAAGTACATGAAGCTGCAGCATGTCAACAAGAGGCTGGAG 1008

Qy    901 cggcgagttattgactggtatcagcacctgcagatcaacaagaagatgaccaacgaggta 960
          |||||||||||||||||||| ||||| || ||||||||||||||||||||||||||||||
Db   1009 CGGCGAGTTATTGACTGGTACCAGCATCTTCAGATCAACAAGAAGATGACCAACGAGGTA 1068

Qy    961 gccatcttacagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctg 1020
          |||||||| |||||| |||||||||||||||||| || || |||||||| ||||||||||
Db   1069 GCCATCTTGCAGCACCTGCCTGAGCGGCTGCGGGCGGAGGTTGCTGTGTCCGTGCACCTG 1128

Qy   1021 tccactctgagccgggtgcagatctttcagaactgtgaggccagcctgctggaggagctg 1080
          || || |||||||| || |||||||| |||||||||| |||||||||||||| ||||||
Db   1129 TCTACCCTGAGCCGAGTACAGATCTTCCAGAACTGTGAAGCCAGCCTGCTGGAAGAGCTG 1188

Qy   1081 gtgctgaagctgcagccccagacctactcaccaggtgaatatgtatgccgcaaaggagac 1140
          |||||||||| |||||||||||||||||| ||||| |||||||| |||||||||| |||
Db   1189 GTGCTGAAGCTACAGCCCCAGACCTACTCGCCAGGCGAATATGTGTGCCGCAAAGGGGAC 1248

Qy   1141 attggccaagagatgtacatcatccgagagggtcaactggccgtggtggcagatgatggt 1200
          ||||||| |||||||||||||||| ||||| || ||||| |||||||| |||||||||
Db   1249 ATTGGCCGAGAGATGTACATCATCCGTGAGGGCCAGCTGGCTGTGGTGGCCGATGATGGT 1308

Qy   1201 atcacacagtatgctgtgctcggtgcagggctctactttggggagatcagcatcatcaac 1260
```

```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 186609)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-290F24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 186609)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 9, 2002 this sequence version replaced gi:18483561.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5842
                Center clone name: 290_F_24
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   186609: contig of 186609 bp in length.
FEATURES              Location/Qualifiers
     source           1. .186609
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="11"
                      /map="11"
                      /clone="RP11-290F24"
                      /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    60812 a  39184 c  37829 g  48784 t
ORIGIN
```

```
  Query Match           37.8%;  Score 652.6;  DB 2;  Length 186609;
  Best Local Similarity  99.4%;  Pred. No. 6.8e-142;
  Matches  655;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy    261 cttccacacaggattcttggaacagggcatcctggtggtggacaagggtaggatctcgag 320
          | ||  | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 156294 CCTCTCCCCAGGATTCTTGGAACAGGGCATCCTGGTGGTGGACAAGGGTAGGATCTCGAG 156353

Qy    321 tcgctacgttcgcacctggagtttcttcttggacctggcttccctgatgcccacagatgt 380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 156354 TCGCTACGTTCGCACCTGGAGTTTCTTCTTGGACCTGGCTTCCCTGATGCCCACAGATGT 156413

Qy    381 ggtctacgtgcggctgggcccgcacacacccaccctgaggctgaaccgctttctccgcgc 440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 156414 GGTCTACGTGCGGCTGGGCCCGCACACACCCACCCTGAGGCTGAACCGCTTTCTCCGCGC 156473

Qy    441 gccccgcctcttcgaggccttcgaccgcacagagacccgcacagcttacccaaatgcctt 500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 156474 GCCCCGCCTCTTCGAGGCCTTCGACCGCACAGAGACCCGCACAGCTTACCCAAATGCCTT 156533

Qy    501 tcgcattgccaagctgatgctttacatttttgtcgtcatccattggaacagctgcctata 560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 156534 TCGCATTGCCAAGCTGATGCTTTACATTTTTGTCGTCATCCATTGGAACAGCTGCCTATA 156593

Qy    561 ctttgccctatcccggtacctgggcttcgggcgtgacgcatgggtgtacccggaccccgc 620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 156594 CTTTGCCCTATCCCGGTACCTGGGCTTCGGGCGTGACGCATGGGTGTACCCGGACCCCGC 156653

Qy    621 gcagcctggctttgagcgcctgcggcgccagtacctctatagcttttacttctccacgct 680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 156654 GCAGCCTGGCTTTGAGCGCCTGCGGCGCCAGTACCTCTATAGCTTTTACTTCTCCACGCT 156713

Qy    681 gatactgactacagtgggcgatacaccgccgccagccagggaagaagagtacctcttcat 740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 156714 GATACTGACTACAGTGGGCGATACACCGCCGCCAGCCAGGGAAGAAGAGTACCTCTTCAT 156773

Qy    741 ggtgggcgacttcctgctggccgtcatgggtttcgccaccatcatgggtagcatgagctc 800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 156774 GGTGGGCGACTTCCTGCTGGCCGTCATGGGTTTCGCCACCATCATGGGTAGCATGAGCTC 156833

Qy    801 tgtcatctacaacatgaacactgcagatgcggctttctacccagatcatgcactggtgaa 860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 156834 TGTCATCTACAACATGAACACTGCAGATGCGGCTTTCTACCCAGATCATGCACTGGTGAA 156893

Qy    861 gaagtacatgaagctgcagcacgtcaaccgcaagctggagcggcgagttattgactggt 919
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 156894 GAAGTACATGAAGCTGCAGCACGTCAACCGCAAGCTGGAGCGGCGAGTTATTGACTGGT 156952
```

```
RESULT   9
BTCNGCHAN
LOCUS       BTCNGCHAN               2641 bp    mRNA    linear   MAM 20-JUL-1995
DEFINITION  B.taurus mRNA for alpha subunit of cone and testis CNG channel.
ACCESSION   X89600
VERSION     X89600.1  GI:908823
KEYWORDS    alpha-subunit; CNG channel.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 2641)
  AUTHORS   Weyand,I., Godde,M., Frings,S., Weiner,J., Muller,F.,
            Altenhofen,W., Hatt,H. and Kaupp,U.B.
  TITLE     Cloning and functional expression of a cyclic-nucleotide-gated
            channel from mammalian sperm
  JOURNAL   Nature 368 (6474), 859-863 (1994)
  MEDLINE   94211295
REFERENCE   2  (bases 1 to 2641)
  AUTHORS   Weyand,I.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUL-1995) I. Weyand, Forschungszentrum Julich,
            Institut fur Biologische Informationsverarbeitung, Postfach 1913,
            52425 Julich, FRG
FEATURES              Location/Qualifiers
     source           1. .2641
                      /organism="Bos taurus"
                      /db_xref="taxon:9913"
     CDS              151. .2271
                      /codon_start=1
                      /product="alpha subunit of CNG-channel expressed in bovine
                      testis and retinal cone"
```

```
            *    45992    53180: contig of 7189 bp in length
            *    53181 53280: gap of      100 bp
            *    53281    62817: contig of 9537 bp in length
            *    62818 62917: gap of      100 bp
            *    62918    70247: contig of 7330 bp in length
            *    70248 70347: gap of      100 bp
            *    70348    78754: contig of 8407 bp in length
            *    78755 78854: gap of      100 bp
            *    78855    86840: contig of 7986 bp in length
            *    86841 86940: gap of      100 bp
            *    86941    98610: contig of 11670 bp in length
            *    98611 98710: gap of      100 bp
            *    98711   111069: contig of 12359 bp in length
            *   111070 111169: gap of      100 bp
            *   111170   122730: contig of 11561 bp in length
            *   122731 122830: gap of      100 bp
            *   122831   135722: contig of 12892 bp in length
            *   135723 135822: gap of      100 bp
            *   135823   150963: contig of 15141 bp in length
            *   150964 151063: gap of      100 bp
            *   151064   164434: contig of 13371 bp in length.
FEATURES             Location/Qualifiers
     source          1. .164434
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-658A5"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1416
                     /note="assembly_fragment"
     misc_feature    1517. .3580
                     /note="assembly_fragment"
     misc_feature    3681. .4383
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    4484. .5844
                     /note="assembly_fragment"
     misc_feature    5945. .7574
                     /note="assembly_fragment"
     misc_feature    7675. .9988
                     /note="assembly_fragment"
     misc_feature    10089. .13426
                     /note="assembly_fragment"
     misc_feature    13527. .15573
                     /note="assembly_fragment"
     misc_feature    15674. .18974
                     /note="assembly_fragment"
     misc_feature    19075. .21809
                     /note="assembly_fragment"
     misc_feature    21910. .25085
                     /note="assembly_fragment"
     misc_feature    25186. .29425
                     /note="assembly_fragment"
     misc_feature    29526. .34133
                     /note="assembly_fragment"
     misc_feature    34234. .40649
                     /note="assembly_fragment"
     misc_feature    40750. .45891
                     /note="assembly_fragment"
     misc_feature    45992. .53180
                     /note="assembly_fragment"
     misc_feature    53281. .62817
                     /note="assembly_fragment"
     misc_feature    62918. .70247
                     /note="assembly_fragment"
     misc_feature    70348. .78754
                     /note="assembly_fragment"
     misc_feature    78855. .86840
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    86941. .98610
                     /note="assembly_fragment"
     misc_feature    98711. .111069
                     /note="assembly_fragment"
     misc_feature    111170. .122730
                     /note="assembly_fragment"
     misc_feature    122831. .135722
                     /note="assembly_fragment"
     misc_feature    135823. .150963
                     /note="assembly_fragment"
     misc_feature    151064. .164434
                     /note="assembly_fragment"
BASE COUNT    50098 a  34031 c  34547 g  43240 t   2518 others
ORIGIN


  Query Match             37.8%;  Score 652.6;  DB 2;  Length 164434;
  Best Local Similarity   99.4%;  Pred. No. 6.8e-142;
  Matches  655;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy     261 cttccacacaggattcttggaacagggcatcctggtggtggacaagggtaggatctcgag 320
           | ||  | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 157488 CCTCTCCCCAGGATTCTTGGAACAGGGCATCCTGGTGGTGGACAAGGGTAGGATCTCGAG 157547

Qy     321 tcgctacgttcgcacctggagtttcttcttggacctggcttccctgatgcccacagatgt 380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 157548 TCGCTACGTTCGCACCTGGAGTTTCTTCTTGGACCTGGCTTCCCTGATGCCCACAGATGT 157607

Qy     381 ggtctacgtgcggctgggcccgcacacacccaccctgaggctgaaccgctttctccgcgc 440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 157608 GGTCTACGTGCGGCTGGGCCCGCACACACCCACCCTGAGGCTGAACCGCTTTCTCCGCGC 157667

Qy     441 gccccgcctcttcgaggccttcgaccgcacagagacccgcacagcttacccaaatgcctt 500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 157668 GCCCCGCCTCTTCGAGGCCTTCGACCGCACAGAGACCCGCACAGCTTACCCAAATGCCTT 157727

Qy     501 tcgcattgccaagctgatgctttacatttttgtcgtcatccattggaacagctgcctata 560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 157728 TCGCATTGCCAAGCTGATGCTTTACATTTTTGTCGTCATCCATTGGAACAGCTGCCTATA 157787

Qy     561 ctttgccctatcccggtacctgggcttcgggcgtgacgcatgggtgtacccggaccccgc 620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 157788 CTTTGCCCTATCCCGGTACCTGGGCTTCGGGCGTGACGCATGGGTGTACCCGGACCCCGC 157847

Qy     621 gcagcctggctttgagcgcctgcggcgccagtacctctatagcttttacttctccacgct 680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 157848 GCAGCCTGGCTTTGAGCGCCTGCGGCGCCAGTACCTCTATAGCTTTTACTTCTCCACGCT 157907

Qy     681 gatactgactacagtgggcgatacaccgccgccagccagggaagaagagtacctcttcat 740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 157908 GATACTGACTACAGTGGGCGATACACCGCCGCCAGCCAGGGAAGAAGAGTACCTCTTCAT 157967

Qy     741 ggtgggcgacttcctgctggccgtcatgggtttcgccaccatcatgggtagcatgagctc 800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 157968 GGTGGGCGACTTCCTGCTGGCCGTCATGGGTTTCGCCACCATCATGGGTAGCATGAGCTC 158027

Qy     801 tgtcatctacaacatgaacactgcagatgcggctttctacccagatcatgcactggtgaa 860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 158028 TGTCATCTACAACATGAACACTGCAGATGCGGCTTTCTACCCAGATCATGCACTGGTGAA 158087

Qy     861 gaagtacatgaagctgcagcacgtcaaccgcaagctggagcggcgagttattgactggt 919
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 158088 GAAGTACATGAAGCTGCAGCACGTCAACCGCAAGCTGGAGCGGCGAGTTATTGACTGGT 158146


RESULT   8
AC022762
LOCUS       AC022762              186609 bp    DNA     linear   HTG 09-FEB-2002
DEFINITION  Homo sapiens chromosome 11 clone RP11-290F24 map 11, *** SEQUENCING
            IN PROGRESS ***, 1 ordered pieces.
ACCESSION   AC022762
VERSION     AC022762.6  GI:18642872
KEYWORDS    HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
```

```
          |||||||||||||||||||| |||||||||||||||||||||||||||||| |||||||||
Db   1309 GTCACACAGTATGCTGTGCTTGGTGCAGGGCTCTACTTTGGGGAGATCAGTATCATCAAC 1368

Qy   1261 atcaaagggaacatgtctgggaaccgccgcacagccaacatcaagagcctaggttattca 1320
          |||||||||||||||||| ||||| || ||||||||||||||||||||||||||||||||
Db   1369 ATCAAAGGGAACATGTCTGGAAACCGACGAACAGCCAACATCAAGAGCCTAGGTTATTCA 1428

Qy   1321 gacctattctgcctgagcaaggaggacctgcgggaggtgctgagcgagtatccacaagca 1380
          ||||| |||||||| |||||||||| |||||||||| ||||| ||||| ||||| ||
Db   1429 GACCTGTTCTGCCTCAGCAAGGAGGATCTGCGGGAGGTACTGAGTGAGTACCCACAGGCC 1488

Qy   1381 cagaccatcatggaggagaaaggacgtgagatcctgctgaaaatgaacaagttggacgtg 1440
          ||| |  |||||||||||||| || || || ||| |||| |||||||| |||||||| |||
Db   1489 CAGGCGGTCATGGAGGAGAAGGGCCGAGAAATCTTGCTCAAAATGAATAAGTTGGATGTG 1548

Qy   1441 aatgctgaggcagctgagatcgccctgcaggaggccacagagtcccggctacgaggccta 1500
          |||||||||||||||||||||||||| |||||||||||||||| |||||   ||||||
Db   1549 AATGCTGAGGCAGCTGAGATCGCCCTCCAGGAGGCCACAGAGTCTCGGCTCAAAGGCCTC 1608

Qy   1501 gaccagcagctggatgatctacagaccaagtttgctcgcctcctggctgagctggagtcc 1560
          |||||||||| |||||||| |||||||||||||||||||| ||||||||||||||||||
Db   1609 GACCAGCAGCTTGATGATCTGCAGACCAAGTTTGCTCGCCTACTGGCTGAGCTGGAGTCC 1668

Qy   1561 agcgcacttaagattgcttaccgcattgaacggctggagtggcagactcgagagtggcca 1620
          || ||||| ||||| |||||||||||| ||| ||||||||||||||||||||||||||||
Db   1669 AGTGCACTGAAGATAGCTTACCGCATCGAAAGGCTGGAGTGGCAGACTCGAGAGTGGCCA 1728

Qy   1621 atgcccgaggacctggctgaggctgatgacgagggtgagcctgaggagggaacttccaaa 1680
          ||||| |||||| ||| |||||||||||| |||| |||||||| || || || |||||
Db   1729 ATGCCAGAGGACATGGGTGAGGCTGATGATGAGGCTGAGCCTGGAGAAGGGACGTCCAAG 1788

Qy   1681 gatgaagagggcagggccagccaggagggaccccccaggtccagagtga 1728
          |||| |||||| | |||  |||||| ||||||| ||||   ||| |||
Db   1789 GATGGAGAGGGAAAGGCTGGCCAGGCGGGACCCTCAGGCATAGAATGA 1836


RESULT   6
AC021935
LOCUS       AC021935              162965 bp    DNA     linear   HTG 03-MAR-2001
DEFINITION  Homo sapiens chromosome 11 clone RP11-645I8 map 11, WORKING DRAFT
            SEQUENCE, 5 unordered pieces.
ACCESSION   AC021935
VERSION     AC021935.3  GI:13184160
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 162965)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-645I8
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 162965)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 2, 2001 this sequence version replaced gi:7652890.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5277
                Center clone name: 645_I_8
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 40% of reads
                Sequencing vector: Plasmid; n/a; 60% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 161664 bases at least Q40
                Consensus quality: 161989 bases at least Q30
                Consensus quality: 162229 bases at least Q20
                Insert size: 158000; agarose-fp
                Insert size: 162565; sum-of-contigs
                Quality coverage: 12.2 in Q20 bases; agarose-fp
                Quality coverage: 11.9 in Q20.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1     5772: contig of 5772 bp in length
            *      5773  5872: gap of      100 bp
            *      5873   114225: contig of 108353 bp in length
            *    114226  114325: gap of      100 bp
            *    114326   119496: contig of 5171 bp in length
            *    119497  119596: gap of      100 bp
            *    119597   161781: contig of 42185 bp in length
            *    161782  161881: gap of      100 bp
            *    161882   162965: contig of 1084 bp in length.
FEATURES             Location/Qualifiers
     source          1. .162965
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-645I8"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .5772
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    5873. .114225
                     /note="assembly_fragment"
     misc_feature    114326. .119496
                     /note="assembly_fragment"
     misc_feature    119597. .161781
                     /note="assembly_fragment"
     misc_feature    161882. .162965
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    54235 a  33886 c  32984 g  41460 t    400 others
ORIGIN


  Query Match             37.8%;  Score 652.6;  DB 2;  Length 162965;
  Best Local Similarity   99.4%;  Pred. No. 6.8e-142;
  Matches  655;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy    261 cttccacacaggattcttggaacagggcatcctggtggtggacaagggtaggatctcgag 320
          | ||  | ||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db 150515 CCTCTCCCCAGGATTCTTGGAACAGGGCATCCTGGTGGTGGACAAGGGTAGGATCTCGAG 150574

Qy    321 tcgctacgttcgcacctggagtttcttcttggacctggcttccctgatgcccacagatgt 380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 150575 TCGCTACGTTCGCACCTGGAGTTTCTTCTTGGACCTGGCTTCCCTGATGCCCACAGATGT 150634

Qy    381 ggtctacgtgcggctgggcccgcacacacccacccctgaggctgaaccgctttctccgcgc 440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 150635 GGTCTACGTGCGGCTGGGCCCGCACACACCCACCCTGAGGCTGAACCGCTTTCTCCGCGC 150694

Qy    441 gccccgcctcttcgaggccttcgaccgcacagagacccgcacagcttacccaaatgcctt 500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 150695 GCCCCGCCTCTTCGAGGCCTTCGACCGCACAGAGACCCGCACAGCTTACCCAAATGCCTT 150754

Qy    501 tcgcattgccaagctgatgctttacatttttgtcgtcatccattggaacagctgcctata 560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 150755 TCGCATTGCCAAGCTGATGCTTTACATTTTTGTCGTCATCCATTGGAACAGCTGCCTATA 150814

Qy    561 ctttgccctatcccggtacctgggcttcgggcgtgacgcatgggtgtacccggaccccgc 620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 150815 CTTTGCCCTATCCCGGTACCTGGGCTTCGGGCGTGACGCATGGGTGTACCCGGACCCCGC 150874

Qy    621 gcagcctggctttgagcgcctgcggcgccagtacctctatagcttttacttctccacgct 680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 150875 GCAGCCTGGCTTTGAGCGCCTGCGGCGCCAGTACCTCTATAGCTTTTACTTCTCCACGCT 150934

Qy    681 gatactgactacagtgggcgatacaccgccgccagccagggaagaagagtacctcttcat 740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 150935 GATACTGACTACAGTGGGCGATACACCGCCGCCAGCCAGGGAAGAAGAGTACCTCTTCAT 150994

Qy    741 ggtgggcgacttcctgctggccgtcatgggtttcgccaccatcatgggtagcatgagctc 800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 150995 GGTGGGCGACTTCCTGCTGGCCGTCATGGGTTTCGCCACCATCATGGGTAGCATGAGCTC 151054

Qy    801 tgtcatctacaacatgaacactgcagatgcggctttctacccagatcatgcactggtgaa 860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 151055 TGTCATCTACAACATGAACACTGCAGATGCGGCTTTCTACCCAGATCATGCACTGGTGAA 151114

Qy    861 gaagtacatgaagctgcagcacgtcaaccgcaagctggagcggcgagttattgactggt 919
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 151115 GAAGTACATGAAGCTGCAGCACGTCAACCGCAAGCTGGAGCGGCGAGTTATTGACTGGT 151173
```

```
RESULT   7
AC036216
LOCUS       AC036216              164434 bp    DNA     linear   HTG 04-MAY-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-658A5 map 11, WORKING DRAFT
            SEQUENCE, 26 unordered pieces.
ACCESSION   AC036216
VERSION     AC036216.2  GI:7705203
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 164434)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-658A5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 164434)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 4, 2000 this sequence version replaced gi:7523886.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L9404
                Center clone name: 658_A_5
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 149016 bases at least Q40
                Consensus quality: 157301 bases at least Q30
                Consensus quality: 160298 bases at least Q20
                Insert size: 157000; agarose-fp
                Insert size: 161934; sum-of-contigs
                Quality coverage: 3.7 in Q20 bases; agarose-fp
                Quality coverage: 3.6 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1416: contig of 1416 bp in length
            *     1417  1516: gap of      100 bp
            *     1517     3580: contig of 2064 bp in length
            *     3581  3680: gap of      100 bp
            *     3681     4383: contig of 703 bp in length
            *     4384  4483: gap of      100 bp
            *     4484     5844: contig of 1361 bp in length
            *     5845  5944: gap of      100 bp
            *     5945     7574: contig of 1630 bp in length
            *     7575  7674: gap of      100 bp
            *     7675     9988: contig of 2314 bp in length
            *     9989 10088: gap of      100 bp
            *    10089    13426: contig of 3338 bp in length
            *    13427 13526: gap of      100 bp
            *    13527    15573: contig of 2047 bp in length
            *    15574 15673: gap of      100 bp
            *    15674    18974: contig of 3301 bp in length
            *    18975 19074: gap of      100 bp
            *    19075    21809: contig of 2735 bp in length
            *    21810 21909: gap of      100 bp
            *    21910    25085: contig of 3176 bp in length
            *    25086 25185: gap of      100 bp
            *    25186    29425: contig of 4240 bp in length
            *    29426 29525: gap of      100 bp
            *    29526    34133: contig of 4608 bp in length
            *    34134 34233: gap of      100 bp
            *    34234    40649: contig of 6416 bp in length
            *    40650 40749: gap of      100 bp
            *    40750    45891: contig of 5142 bp in length
            *    45892 45991: gap of      100 bp
```

```
                     /protein_id="CAA61759.1"
                     /db_xref="GI:908824"
                     /db_xref="SWISS-PROT:Q29441"
                     /translation="MAKISTQYSHPTRTHPSVRTMDRDLDCIENGLSRTHLPCEETSS
                     ELQEGIAMETRGLAESRQSSFTSQGPTRLSRLIISLRAWSARHLHQEDQRPDSFLERF
                     RGAELQEVSSRESHVQFNVGSQEPPDRGRSAWPLARNNTNTCNNSEKDDKAKKEEKEK
                     KEEKKENPKKEEKKKDSVVMDPSSNMYYHWLTVIAVPVFYNWCLLVCRACFDELQSEH
                     LMLWLVLDYSADILYGMDMLVRARTGFLEQGLMVMDASRLWKHYTQTLHFKLDVLSLV
                     PTDLAYFKLGMNYPELRFNRLLKLARLFEFFDRTETRTNYPNMFRIGNLVLYILIIIH
                     WNACIYFAISKFIGFGTDSWVYPNVSNPEYGRLSRKYIYSLYWSTLTLTTIGETPPPV
                     KDEEYLFVVIDFLVGVLIFATIVGNVGSMISNMNASRAEFQAKIDSIKQYMQFRKVTK
                     DLETRVIRWFDYLWANKKTVDEKEVLKSLPDKLKAEIAINVHLDTLRKVRIFQDCEAG
                     LLVELVLKLRPAVFSPGDYICKKGDIGREMYIIKEGKLAVVAEDGITQFVVLGDGSYF
                     GEISILNIKGSKSGNRRTANIRSIGYSDLFCLSKDDLMEALTEYPEAKKALEEKGRQI
                     LMKDNLIDEELAKAGADPKDIEEKVEHLETSLDSLQTRFARLLAEYNATQMKVKQRLS
                     QLESQVKMGLPPDGDAPQTEASQP"
BASE COUNT      623 a    727 c    762 g    529 t
ORIGIN


  Query Match           35.7%;  Score 617.4;  DB 4;  Length 2641;
  Best Local Similarity  62.3%;  Pred. No. 1.1e-133;
  Matches  988;  Conservative    0;  Mismatches  596;  Indels    3;  Gaps    1;

Qy      75 tgtcctggacccatctggggattactactactggtggctgaacacaatggtcttcccagt 134
           ||| ||||||| ||  |  |    ||||||  |||||||     || | | | || ||
Db     684 TGTGATGGACCCTTCCAGCAACATGTACTACCACTGGCTGACTGTCATCGCCGTGCCTGT 743

Qy     135 catgtataacctcatcatcctcgtgtgcagagcctgcttccccgacttgcagcacggtta 194
           | | || |||       | ||||||||||| ||||| |||   ||  | |||  ||   |
Db     744 CTTCTACAACTGGTGTCTGCTCGTGTGCAGGGCCTGTTTCGATGAGCTCCAGTCCGAGCA 803

Qy     195 tctggtggcctggttggtgctggactacacgagtgacctgctatacctactagacatggt 254
           ||| ||   ||| |||| |||||||||| |    ||| | || ||      | || ||| |
Db     804 CCTGATGCTTTGGCTGGTCCTGGACTACTCAGCAGACATCCTCTATGGCATGGATATGCT 863

Qy     255 ggtgcgcttccacacaggattcttggaacagggcatcctggtggtggacaagggtaggat 314
           ||| ||   ||  ||||| ||| |||| || ||| |  |||| |||||    |  || |
Db     864 GGTCCGAGCCCGGACAGGCTTCCTGGAGCAAGGCCTGATGGTCATGGACGCCAGCCGGCT 923

Qy     315 ctcgagtcgctacgttcgcacctggagtttcttcttggacctggcttccctgatgcccac 374
           | ||  | ||||   |  |||| |   |||    ||||| ||   |||||| |||||||
Db     924 GTGGAAGCACTACACACAGACCTTGCACTTCAAGCTGGACGTGTTGTCCCTGGTGCCCAC 983

Qy     375 agatgtggtctacgtgcggctgggcccgcacacacccaccctgaggctgaaccgctttct 434
           |||  |||  ||  |  |||||||  | ||   ||    ||||||  | |||||| | ||
Db     984 AGACCTGGCTTATTTTAAGCTGGGCATGAACTACCCAGAACTGAGGTTCAACCGCCTCCT 1043

Qy     435 ccgcgcgccccgcctcttcgaggccttcgaccgcacagagacccgcacagcttacccaaa 494
           |  |||| |||||||   |||  |||||||| |||||  |  ||    |||||  ||
Db    1044 GAAGTTGGCCCGGCTCTTCGAATTCTTTGACCGCACGGAGACAAGGACCAACTACCCCAA 1103

Qy     495 tgccttcgcattgccaagctgatgctttacatttttgtcgtcatccattggaacagctg 554
           ||  | || |  ||  || |  | |||||  |  || |||||||  |||||   |||
Db    1104 CATGTTCAGGATCGGGAACTTGGTCTTGTACATCCTCATCATCATCCACTGGAATGCCTG 1163

Qy     555 cctatactttgccctatcccggtacctgggcttcgggcgtgacgcatgggtgtacccgga 614
           | | |||||||||| | |||  || | | || ||||||   ||| | ||||| |||||  |
Db    1164 CATCTACTTTGCCATTTCCAAGTTCATTGGTTTCGGGACCGACTCCTGGGTCTACCCAAA 1223

Qy     615 ccccgcgcagcctggctttgagcgcctgcggcgccagtacctctatagcttttacttctc 674
           |  | |  | || |  | || |||||      |  ||||| | || ||  | ||||  ||
Db    1224 CGTCTCCAACCCAGAGTATGGGCGCCTCTCCAGAAAGTACATTTACAGTCTCTACTGGTC 1283

Qy     675 cacgctgatactgactacagtgggcgatacaccgccgccagccagggaagaagagtacct 734
           |||  |||  ||||| ||  | || || || ||||| || |  |  || || || || ||
Db    1284 CACCTTGACCCTGACCACCATTGGGGAGACCCCGCCCCCCGTAAAAGACGAGGAATATCT 1343

Qy     735 cttcatggtgggcgacttcctgctggccgtcatgggtttcgccaccatcatgggtagcat 794
           |||  ||||   |||||||||| ||| |||| ||  ||| |||||||| |||| | | |
Db    1344 CTTTGTGGTCATCGACTTCCTGGTGGGCGTCCTGATTTTTGCCACCATCGTGGGCAACGT 1403

Qy     795 gagctctgtcatctacaacatgaacactgcagatgcggctttctacccagatcatgcact 854
           | ||||  | ||||  |||||||| || ||  || |  ||| |  |  |    |
Db    1404 GGGCTCCATGATCTCAAACATGAATGCTTCACGGGCCGAGTTCCAGGCCAAGATCGATTC 1463

Qy     855 ggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggagcggcgagttattga 914
           | ||| |||||||| || | |   | || | |    |  |||||   || || ||
Db    1464 CATCAAGCAGTACATGCAGTTCCGCAAGGTGACCAAGGACTTGGAGACACGGGTGATCCG 1523

Qy     915 ctggtatcagcacctgcagatcaacaagaagatgaccaacgaggtagccatcttacagca 974
           |||||   |  |||||  |  ||| |||||||      | |||   |   |  |  ||
Db    1524 CTGGTTCGACTACCTGTGGGCCAATAAGAAGACAGTGGATGAGAAGGAGGTGCTCAAGAG 1583

Qy     975 cttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctgtccactctgagccg 1034
           | | || ||   ||||  ||| ||  | ||  |    ||||||||  ||| |||||
Db    1584 CCTCCCCGACAAGCTGAAGGCCGAGATCGCCATCAACGTGCACCTGGACACCCTGAGGAA 1643

Qy    1035 ggtgcagatctttcagaactgtgaggccagcctgctggaggagctggtgctgaagctgca 1094
           ||| |  ||||| ||| |||| |||||  | |||||| ||||||||||||||||||||||
Db    1644 GGTCCGAATCTTCCAGGACTGCGAGGCGGGGCTGCTGGTGGAGCTGGTGCTGAAGCTGCG 1703

Qy    1095 gccccagacctactcaccaggtgaatatgtatgccgcaaaggagacattggccaagagat 1154
           |||       | |   || || || ||  | |||   || || ||||| ||    |||||
Db    1704 GCCGGCAGTGTTCAGCCCCGGGGACTACATCTGCAAGAAGGGGGACATCGGGAGGGAGAT 1763

Qy    1155 gtacatcatccgagagggtcaactggccgtggtggcagatgatggtatcacacagtatgc 1214
           ||||||  |||   |||||  | ||||||||||||||| || || || ||||| ||||  |
Db    1764 GTACATTATCAAGGAGGGCAAGCTGGCCGTGGTGGCCGAGGACGGCATCACCCAGTTCGT 1823

Qy    1215 tgtgctcggtgcagggctctactttggggagatcagcatcatcaacatcaaagggaacat 1274
           || ||||| |  |||   || || |||||||||||||| || ||| |||||||| |||| ||
Db    1824 GGTCCTCGGCGACGGGAGTTATTTCGGGGAGATCAGCATCTTGAACATCAAGGGGAGCAA 1883

Qy    1275 gtctgggaaccgccgcacagccaacatcaagagcctaggttattcagacctattctgcct 1334
           ||| |||||||||||||| |||||||||| |||| | || || || ||||| ||||||||
Db    1884 GTCCGGGAACCGCCGCACGGCCAACATCAGGAGCATCGGCTACTCGGACCTGTTCTGCCT 1943

Qy    1335 gagcaaggaggacctgcgggaggtgctgagcgagtatccacaagcacagaccatcatgga 1394
           ||||||| ||||||  ||||| ||| | |||||| ||  | ||  |||      ||||
Db    1944 CTCCAAGGACGACCTGATGGAGGCGCTCACCGAGTACCCCGAGGCCAAGAAGGCGCTGGA 2003

Qy    1395 ggagaaaggacgtgagatcctgctgaaaatgaacaagttggacgtgaatgctgaggcagc 1454
           |||||||||| ||  |||||||  ||||    |||  | | |||| | |     |    ||
Db    2004 GGAGAAAGGGCGGCAGATCCTCATGAAGGACAACCTGATCGACGAGGAGCTGGCCAAGGC 2063

Qy    1455 tgagatcgccctgcaggaggccacagagtcccggctacgaggcctagaccagcagctgga 1514
           | |   | ||   ||||   ||  |||      | |       ||| ||       |||||
Db    2064 CGGGGCAGACCCCAAGGA---CATCGAGGAGAAGGTGGAGCACCTCGAGACCTCCCTGGA 2120

Qy    1515 tgatctacagaccaagtttgctcgcctcctggctgagctggagtccagcgcacttaagat 1574
              || ||||||| |||||| || ||||||||||||    |  ||| |    | ||| |
Db    2121 CTCCCTGCAGACCAGGTTTGCGCGGCTCCTGGCTGAGTACAACGCCACCCAGATGAAGGT 2180

Qy    1575 tgcttaccgcattgaacggctggagtggcagactcgagagtggccaatgcccgaggacct 1634
              | ||  |    | | ||||  | |||         | | |    ||| || | |
Db    2181 GAAGCAGCGGCTCAGCCAGTTGGAAAGCCAGGTGAAGATGGGCCTCCCGCCTGATGGCGA 2240

Qy    1635 ggctgaggctgatgacgagggtgagcc 1661
           |||  |    ||| |   || ||||
Db    2241 TGCTCCGCAAACTGAGGCCAGTCAGCC 2267


RESULT  10
BTCNG3
LOCUS       BTCNG3                  3998 bp    mRNA    linear   MAM 19-MAY-1994
DEFINITION  B.taurus CNG3 mRNA for cyclic nucleotide-gated channel.
ACCESSION   X76485
VERSION     X76485.1  GI:488728
KEYWORDS    cyclic nucleotide-gated channel.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 3998)
```

```
  AUTHORS   Biel,M., Zong,X., Distler,M., Bosse,E., Klugbauer,N., Murakami,M.,
            Flockerzi,V. and Hofmann,F.
  TITLE     Another member of the cyclic nucleotide-gated channel family,
            expressed in testis, kidney, and heart
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 91 (9), 3505-3509 (1994)
  MEDLINE   94224768
REFERENCE   2  (bases 1 to 3998)
  AUTHORS   Hofmann,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-NOV-1993) F. Hofmann, Inst. f. Pharmakologie Und
            Toxikologie, Technische Univ. Muenchen, Biedersteiner Str. 29,
            80802 Muenchen, FRG
FEATURES             Location/Qualifiers
     source          1. .3998
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /tissue_type="kidney"
                     /clone_lib="plasmid vector pSport (BRL)"
                     /dev_stage="adult"
     gene            273. .2393
                     /gene="CNG3"
     CDS             273. .2393
                     /gene="CNG3"
                     /codon_start=1
                     /product="cyclic nucleotide-gated channel 3"
                     /protein_id="CAA54023.1"
                     /db_xref="GI:488729"
                     /db_xref="SWISS-PROT:Q29441"
                     /translation="MAKISTQYSHPTRTHPSVRTMDRDLDCIENGLSRTHLPCEETSS
                     ELQEGIAMETRGLAESRQSSFTSQGPTRLSRLIISLRAWSARHLHQEDQRPDSFLERF
                     RGAELQEVSSRESHVQFNVGSQEPPDRGRSAWPLARNNTNTCNNSEKDDKAKKEEKEK
                     KEEKKENPKKEEKKKDSVVMDPSSNMYYHWLTVIAVPVFYNWCLLVCRACFDELQSEH
                     LMLWLVLDYSADILYGMDMLVRARTGFLEQGLMVMDASRLWKHYTQTLHFKLDVLSLV
                     PTDLAYFKLGMNYPELRFNRLLKLARLFEFFDRTETRTNYPNMFRIGNLVLYILIIIH
                     WNACIYFAISKFIGFGTDSWVYPNVSNPEYGRLSRKYIYSLYWSTLTLTTIGETPPPV
                     KDEEYLFVVIDFLVGVLIFATIVGNVGSMISNMNASRAEFQAKIDSIKQYMQFRKVTK
                     DLETRVIRWFDYLWANKKTVDEKEVLKSLPDKLKAEIAINVHLDTLRKVRIFQDCEAG
                     LLVELVLKLRPAVFSPGDYICKKGDIGREMYIIKEGKLAVVAEDGITQFVVLGDGSYF
                     GEISILNIKGSKSGNRRTANIRSIGYSDLFCLSKDDLMEALTEYPEAKKALEEKGRQI
                     LMKDNLIDEELAKAGADPKDIEEKVEHLETSLDSLQTRFARLLAEYNATQMKVKQRLS
                     QLESQVKMGLPPDGDAPQTEASQP"
BASE COUNT      997 a   1054 c   1050 g    897 t
ORIGIN
```

```
  Query Match             35.7%;  Score 617.4;  DB 4;  Length 3998;
  Best Local Similarity   62.3%;  Pred. No. 1.1e-133;
  Matches  988;  Conservative    0;  Mismatches  596;  Indels    3;  Gaps    1;

Qy      75 tgtcctggacccatctggggattactactactggtggctgaacacaatggtcttcccagt 134
           ||| ||||||| ||  |  |    ||||||   |||||||     || | | | || ||
Db     806 TGTGATGGACCCTTCCAGCAACATGTACTACCACTGGCTGACTGTCATCGCCGTGCCTGT 865

Qy     135 catgtataacctcatcatcctcgtgtgcagagcctgcttccccgacttgcagcacggtta 194
           | | || |||       | |||||||||| ||||| |||   ||  | |||  ||   |
Db     866 CTTCTACAACTGGTGTCTGCTCGTGTGCAGGGCCTGTTTCGATGAGCTCCAGTCCGAGCA 925

Qy     195 tctggtggcctggttggtgctggactacacgagtgacctgctatacctactagacatggt 254
           ||| ||   ||| |||| ||||||||| |    ||| | || ||     | || ||| |
Db     926 CCTGATGCTTTGGCTGGTCCTGGACTACTCAGCAGACATCCTCTATGGCATGGATATGCT 985

Qy     255 ggtgcgcttccacacaggattcttggaacagggcatcctggtggtggacaagggtaggat 314
           ||| ||   ||  ||||| ||| |||| || ||| |  ||||  |||||    |  || |
Db     986 GGTCCGAGCCCGGACAGGCTTCCTGGAGCAAGGCCTGATGGTCATGGACGCCAGCCGGCT 1045

Qy     315 ctcgagtcgctacgttcgcacctggagtttcttcttggacctggcttccctgatgcccac 374
           | ||  | ||||   |  |||| |   |||    ||||| ||   |||||| |||||||
Db    1046 GTGGAAGCACTACACACAGACCTTGCACTTCAAGCTGGACGTGTTGTCCCTGGTGCCCAC 1105

Qy     375 agatgtggtctacgtgcggctgggcccgcacacacccaccctgaggctgaaccgctttct 434
           |||  |||  ||  |   |||||||  | ||   ||    |||||| | |||||| | ||
Db    1106 AGACCTGGCTTATTTTAAGCTGGGCATGAACTACCCAGAACTGAGGTTCAACCGCCTCCT 1165

Qy     435 ccgcgcgccccgcctcttcgaggccttcgaccgcacagagacccgcacagcttacccaaa 494
                 | |||| ||||||||   ||| |||||||| |||||  | ||   |||||  ||
Db    1166 GAAGTTGGCCCGGCTCTTCGAATTCTTTGACCGCACGGAGACAAGGACCAACTACCCCAA 1225

Qy     495 tgcctttcgcattgccaagctgatgctttacatttttgtcgtcatccattggaacagctg 554
              ||  | ||  |  ||  || ||  | |||||  |  || |||||||  |||||   |||
Db    1226 CATGTTCAGGATCGGGAACTTGGTCTTGTACATCCTCATCATCATCCACTGGAATGCCTG 1285

Qy     555 cctatactttgccctatcccggtacctgggcttcgggcgtgacgcatgggtgtacccgga 614
           | | |||||||| | |||  || | | || ||||||    ||| | ||||| |||||  |
Db    1286 CATCTACTTTGCCATTTCCAAGTTCATTGGTTTCGGGACCGACTCCTGGGTCTACCCAAA 1345

Qy     615 ccccgcgcagcctggctttgagcgcctgcggcgccagtacctctatagcttttacttctc 674
           |  | |  | || |  | || ||||||      |  ||||| | || ||  | |||| ||
Db    1346 CGTCTCCAACCCAGAGTATGGGCGCCTCTCCAGAAAGTACATTTACAGTCTCTACTGGTC 1405

Qy     675 cacgctgatactgactacagtgggcgatacaccgccgccagccagggaagaagagtacct 734
           |||  |||  ||||| ||  | || || || ||||| || |  |  || || || || ||
Db    1406 CACCTTGACCCTGACCACCATTGGGGAGACCCCGCCCCCCGTAAAAGACGAGGAATATCT 1465

Qy     735 cttcatggtgggcgacttcctgctggccgtcatgggtttcgccaccatcatgggtagcat 794
           |||  ||||   |||||||||| ||| |||| ||  ||| |||||||| |||| | | |
Db    1466 CTTTGTGGTCATCGACTTCCTGGTGGGCGTCCTGATTTTTGCCACCATCGTGGGCAACGT 1525

Qy     795 gagctctgtcatctacaacatgaacactgcagatgcggctttctacccagatcatgcact 854
           | ||||  | ||||  |||||||| || ||   || |  ||| |  |  |    |
Db    1526 GGGCTCCATGATCTCAAACATGAATGCTTCACGGGCCGAGTTCCAGGCCAAGATCGATTC 1585

Qy     855 ggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggagcggcgagttattga 914
             | ||| |||||||| || | |   | || | |    |  |||||   || || ||
Db    1586 CATCAAGCAGTACATGCAGTTCCGCAAGGTGACCAAGGACTTGGAGACACGGGTGATCCG 1645

Qy     915 ctggtatcagcacctgcagatcaacaagaagatgaccaacgaggtagccatcttacagca 974
           |||||   |  |||||  |  ||| |||||||      | |||   |   |  |  ||
Db    1646 CTGGTTCGACTACCTGTGGGCCAATAAGAAGACAGTGGATGAGAAGGAGGTGCTCAAGAG 1705

Qy     975 cttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctgtccactctgagccg 1034
           | | || ||   ||||  ||| ||  | ||  |    |||||||||  ||| |||||
Db    1706 CCTCCCCGACAAGCTGAAGGCCGAGATCGCCATCAACGTGCACCTGGACACCCTGAGGAA 1765

Qy    1035 ggtgcagatctttcagaactgtgaggccagcctgctggaggagctggtgctgaagctgca 1094
           ||| |  ||||| ||| |||| |||||  | |||||| ||||||||||||||||||||
Db    1766 GGTCCGAATCTTCCAGGACTGCGAGGCGGGGCTGCTGGTGGAGCTGGTGCTGAAGCTGCG 1825

Qy    1095 gccccagacctactcaccaggtgaatatgtatgccgcaaaggagacattggccaagagat 1154
           |||        | |   || || || ||  | |||   || || ||||| ||    |||||
Db    1826 GCCGGCAGTGTTCAGCCCCGGGGACTACATCTGCAAGAAGGGGGACATCGGGAGGGAGAT 1885

Qy    1155 gtacatcatccgagagggtcaactggccgtggtggcagatgatggtatcacacagtatgc 1214
           |||||| |||   |||||  | ||||||||||||||| || || || ||||| |||| |
Db    1886 GTACATTATCAAGGAGGGCAAGCTGGCCGTGGTGGCCGAGGACGGCATCACCCAGTTCGT 1945

Qy    1215 tgtgctcggtgcagggctctactttggggagatcagcatcatcaacatcaaagggaacat 1274
             || ||||| |  |||   || || ||||||||||||||| | |||||||| |||| ||
Db    1946 GGTCCTCGGCGACGGGAGTTATTTCGGGGAGATCAGCATCTTGAACATCAAGGGGAGCAA 2005

Qy    1275 gtctgggaaccgccgcacagccaacatcaagagcctaggttattcagacctattctgcct 1334
           ||| |||||||||||||| |||||||||| |||| | || || || ||||| ||||||||
Db    2006 GTCCGGGAACCGCCGCACGGCCAACATCAGGAGCATCGGCTACTCGGACCTGTTCTGCCT 2065

Qy    1335 gagcaaggaggacctgcgggaggtgctgagcgagtatccacaagcacagaccatcatgga 1394
             ||||||  ||||||  |||||  ||| | ||||||  ||  |  ||  |||      ||||
Db    2066 CTCCAAGGACGACCTGATGGAGGCGCTCACCGAGTACCCCGAGGCCAAGAAGGCGCTGGA 2125

Qy    1395 ggagaaaggacgtgagatcctgctgaaaatgaacaagttggacgtgaatgctgaggcagc 1454
           |||||||||  ||  |||||||  ||||    |||  | | ||||  | |    |     ||
Db    2126 GGAGAAAGGGCGGCAGATCCTCATGAAGGACAACCTGATCGACGAGGAGCTGGCCAAGGC 2185

Qy    1455 tgagatcgccctgcaggaggccacagagtcccggctacgaggcctagaccagcagctgga 1514
           | |   | ||   ||||  .||  |||      | |       |||  ||       |||||
Db    2186 CGGGGCAGACCCCAAGGA---CATCGAGGAGAAGGTGGAGCACCTCGAGACCTCCCTGGA 2242

Qy    1515 tgatctacagaccaagtttgctcgcctcctggctgagctggagtccagcgcacttaagat 1574
```

```
Db   2243 CTCCCTGCAGACCAGGTTTGCGCGGCTCCTGGCTGAGTACAACGCCACCCAGATGAAGGT 2302

Qy   1575 tgcttaccgcattgaacggctggagtggcagactcgagagtggccaatgcccgaggacct 1634

Db   2303 GAAGCAGCGGCTCAGCCAGTTGGAAAGCCAGGTGAAGATGGGCCTCCCGCCTGATGGCGA 2362

Qy   1635 ggctgaggctgatgacgagggtgagcc 1661

Db   2363 TGCTCCGCAAACTGAGGCCAGTCAGCC 2389


RESULT 11
AF065314
LOCUS       AF065314               3486 bp    mRNA    linear   PRI 27-MAY-1998
DEFINITION  Homo sapiens cone photoreceptor cGMP-gated channel alpha subunit
            (CNGA3) mRNA, complete cds.
ACCESSION   AF065314
VERSION     AF065314.1  GI:3153886
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3486)
  AUTHORS   Wissinger,B., Muller,F., Weyand,I., Schuffenhauer,S., Thanos,S.,
            Kaupp,U.B. and Zrenner,E.
  TITLE     Cloning, chromosomal localization and functional expression of the
            gene encoding the alpha-subunit of the cGMP-gated channel in human
            cone photoreceptors
  JOURNAL   Eur. J. Neurosci. 9 (12), 2512-2521 (1997)
  MEDLINE   98176633
REFERENCE   2  (bases 1 to 3486)
  AUTHORS   Wissinger,B., Muller,F., Weyand,I., Schuffenhauer,S., Thanos,S.,
            Kaupp,U.B. and Zrenner,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAY-1998) Molekulargenetisches Labor,
            Universitaets-Augenklinik Abt II, Verfuegungsgebaeude, Auf der
            Morgenstelle 15, Tuebingen D-72076, Germany
FEATURES             Location/Qualifiers
     source          1. .3486
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2q11"
     gene            1. .3486
                     /gene="CNGA3"
     CDS             40. .2124
                     /gene="CNGA3"
                     /codon_start=1
                     /product="cone photoreceptor cGMP-gated channel alpha
                     subunit"
                     /protein_id="AAC17440.1"
                     /db_xref="GI:3153887"
                     /translation="MAKINTQYSHPSRTHLKVKTSDRDLNRAENGLSRAHSSSEETSS
                     VLQPGIAMETRGLADSGQGSFTGQGIARLSRLIFLLRRWAARHVHHQDQGPDSFPDRF
                     RGAELKEVSSQESNAQANVGSQEPADRGRSAWPLAKCNTNTSNNTEEEKKTKKKDAIV
                     VDPSSNLYYRWLTAIALPVFYNWYLLICRACFDELQSEYLMLWLVLDYSADVLYVLDV
                     LVRARTGFLEQGLMVSDTNRLWQHYKTTTQFKLDVLSLVPTDLAYLKVGTNYPEVRFN
                     RLLKFSRLFEFFDRTETRTNYPNMFRIGNLVLYILIIIHWNACIYFAISKFIGFGTDS
                     WVYPNISIPEHGRLSRKYIYSLYWSTLTLTTIGETPPPVKDEEYLFVVVDFLVGVLIF
                     ATIVGNVGSMISNMNASRAEFQAKIDSIKQYMQFRKVTKDLETRVIRWFDYLWANKKT
                     VDEKEVLKSLPDKLKAEIAINVHLDTLKKVRIFQDCEAGLLVELVLKLRPTVFSPGDY
                     ICKKGDIGKEMYIINEGKLAVVADDGVTQFVVLSDGSYFGEISILNIKGSKSGNRRTA
                     NIRSIGYSDLFCLSKDDLMEALTEYPEAKKALEEKGRQILMKDNLIDEELARAGADPK
                     DLEEKVEQLGSSLDTLQTRFARLLAEYNATQMKMKQRLSQLESQVKGGGDKPLADGEV
                     PGDATKTEDKQQ"
BASE COUNT      912 a    866 c    865 g    843 t
ORIGIN


  Query Match           34.0%;  Score 588.2;  DB 9;  Length 3486;
  Best Local Similarity 61.7%;  Pred. No. 7.2e-127;

  Matches  954;  Conservative    0;  Mismatches  588;  Indels    3;  Gaps    1;

Qy     61 aggaagttgctgcctgtcctggacccatctggggattactactactggtggctgaacaca 120

Db    502 AAGAAGGATGCGATCGTGGTGGACCCGTCCAGCAACCTGTACTACCGCTGGCTGACCGCC 561

Qy    121 atggtcttcccagtcatgtataacctcatcatcctcgtgtgcagagcctgcttccccgac 180

Db    562 ATCGCCCTGCCTGTCTTCTATAACTGGTATCTGCTTATTTGCAGGGCCTGTTTCGATGAG 621

Qy    181 ttgcagcacggttatctggtggcctggttggtgctggactacacgagtgacctgctatac 240

Db    622 CTGCAGTCCGAGTACCTGATGCTGTGGCTGGTCCTGGACTACTCGGCAGATGTCCTGTAT 681

Qy    241 ctactagacatggtggtgcgcttccacacaggattcttggaacagggcatcctggtggtg 300

Db    682 GTCTTGGATGTGCTTGTACGAGCTCGGACAGGTTTTCTCGAGCAAGGCTTAATGGTCAGT 741

Qy    301 gacaagggtaggatctcgagtcgctacgttcgcacctggagtttcttcttggacctggct 360

Db    742 GATACCAACAGGCTGTGGCAGCATTACAAGACGACCACGCAGTTCAAGCTGGATGTGTTG 801

Qy    361 tccctgatgcccacagatgtggtctacgtgcggctgggcccgcacacacccaccctgagg 420

Db    802 TCCCTGGTCCCCACCGACCTGGCTTACTTAAAGGTGGGCACAAACTACCCAGAAGTGAGG 861

Qy    421 ctgaaccgctttctccgcgcgccccgcctcttcgaggccttcgaccgcacagagacccgc 480

Db    862 TTCAACCGCCTACTGAAGTTTTCCCGGCTCTTTGAATTCTTTGACCGCACAGAGACAAGG 921

Qy    481 acagcttacccaaatgcctttcgcattgccaagctgatgctttacatttttgtcgtcatc 540

Db    922 ACCAACTACCCCAATATGTTCAGGATTGGGAACTTGGTCTTGTACATTCTCATCATCATC 981

Qy    541 cattggaacagctgcctatactttgccctatcccggtacctgggcttcgggcgtgacgca 600

Db    982 CACTGGAATGCCTGCATCTACTTTGCCATTTCCAAGTTCATTGGTTTTGGGACAGACTCC 1041

Qy    601 tgggtgtacccggaccccgcgcagcctggctttgagcgcctgcggcgccagtacctctat 660

Db   1042 TGGGTCTACCCAAACATCTCAATCCCAGAGCATGGGCGCCTCTCCAGGAAGTACATTTAC 1101

Qy    661 agcttttacttctccacgctgatactgactacagtgggcgatacaccgccgccagccagg 720

Db   1102 AGTCTCTACTGGTCCACCTTGACCCTTACCACCATTGGTGAGACCCCACCCCCCGTGAAA 1161

Qy    721 gaagaagagtacctcttcatggtgggcgacttcctgctggccgtcatgggtttcgccacc 780

Db   1162 GATGAGGAGTATCTCTTTGTGGTCGTAGACTTCTTGGTGGGTGTTCTGATTTTTGCCACC 1221

Qy    781 atcatgggtagcatgagctctgtcatctacaacatgaacactgcagatgcggctttctac 840

Db   1222 ATTGTGGGCAATGTGGGCTCCATGATCTCGAATATGAATGCCTCACGGGCAGAGTTCCAG 1281

Qy    841 ccagatcatgcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggag 900

Db   1282 GCCAAGATTGATTCCATCAAGCAGTACATGCAGTTCCGCAAGGTCACCAAGGACTTGGAG 1341

Qy    901 cggcgagttattgactggtatcagcacctgcagatcaacaagaagatgaccaacgaggta 960

Db   1342 ACGCGGGTTATCCGGTGGTTTGACTACCTGTGGGCCAACAAGAAGACGGTGGATGAGAAG 1401

Qy    961 gccatcttacagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctg 1020

Db   1402 GAGGTGCTCAAGAGCCTCCCAGACAAGCTGAAGGCTGAGATCGCCATCAACGTGCACCTG 1461

Qy   1021 tccactctgagccgggtgcagatctttcagaactgtgaggccagcctgctggaggagctg 1080

Db   1462 GACACGCTGAAGAAGGTTCGCATCTTCCAGGACTGTGAGGCAGGGCTGCTGGTGGAGCTG 1521

Qy   1081 gtgctgaagctgcagccccagacctactcaccaggtgaatatgtatgccgcaaaggagac 1140

Db   1522 GTGCTGAAGCTGCGACCCACTGTGTTCAGCCCTGGGGATTATATCTGCAAGAAGGGAGAT 1581
```

```
Qy   1141 attggccaagagatgtacatcatccgagagggtcaactggccgtggtggcagatgatggt 1200
          ||||| | ||||||||||||||||   |||||  | ||||||||||||||| ||||||||
Db   1582 ATTGGGAAGGAGATGTACATCATCAACGAGGGCAAGCTGGCCGTGGTGGCTGATGATGGG 1641

Qy   1201 atcacacagtatgctgtgctcggtgcagggctctactttggggagatcagcatcatcaac 1260
          |||| ||||  |  || ||| | |  ||   |||||| |||||||||||||||  | |||
Db   1642 GTCACCCAGTTCGTGGTCCTCAGCGATGGCAGCTACTTCGGGGAGATCAGCATTCTGAAC 1701

Qy   1261 atcaaagggaacatgtctgggaaccgccgcacagccaacatcaagagcctaggttattca 1320
          ||||| |||| || ||| |||||||||  | || |||||||||   ||| | || || |||
Db   1702 ATCAAGGGGAGCAAGTCGGGGAACCGCAGGACGGCCAACATCCGCAGCATTGGCTACTCA 1761

Qy   1321 gacctattctgcctgagcaaggaggacctgcgggaggtgctgagcgagtatccacaagca 1380
          ||||| ||||||||    ||||| || ||   |||||  || | |||||| ||  ||||
Db   1762 GACCTGTTCTGCCTCTCAAAGGACGATCTCATGGAGGCCCTCACCGAGTACCCCGAAGCC 1821

Qy   1381 cagaccatcatggaggagaaaggacgtgagatcctgctgaaaatgaacaagttggacgtg 1440
          |||    | ||||||||||||||||  |||||||| |||||    ||||  ||  ||
Db   1822 AAGAAGGCCCTGGAGGAGAAAGGACGGCAGATCCTGATGAAA---GACAACCTGATCGAT 1878

Qy   1441 aatgctgaggcagctgagatcgccctgcaggaggccacagagtcccggctacgaggccta 1500
          | |   |||   | |  |||    |  ||| |   |||      |       ||
Db   1879 GAGGAGCTGGCCAGGGCGGGCGCGGACCCCAAGGACCTTGAGGAGAAAGTGGAGCAGCTG 1938

Qy   1501 gaccagcagctggatgatctacagaccaagtttgctcgcctcctggctgagctggagtcc 1560
          |       |||||    || ||||||| |||||| ||||||||||||||    |  ||
Db   1939 GGGTCCTCCCTGGACACCCTGCAGACCAGGTTTGCACGCCTCCTGGCTGAGTACAACGCC 1998

Qy   1561 agcgcacttaagattgcttaccgcattgaacggctggagtggcag 1605
          | |    | |||||     | ||  |    |  |||||  | |||
Db   1999 ACCCAGATGAAGATGAAGCAGCGTCTCAGCCAACTGGAAAGCCAG 2043
```

```
RESULT 12
MMU243933
LOCUS       MMU243933               2781 bp    mRNA    linear   ROD 25-MAY-2000
DEFINITION  Mus musculus mRNA for cone photoreceptor cyclic nucleotide-gated
            channel alpha-subunit.
ACCESSION   AJ243933
VERSION     AJ243933.1  GI:7688040
KEYWORDS    CNG3 gene; cyclic nucleotide-gated channel alpha-subunit.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2781)
  AUTHORS   Hirano,A.A., Hack,I., Wassle,H. and Duvoisin,R.M.
  TITLE     Cloning and immunocytochemical localization of a cyclic
            nucleotide-gated channel alpha-subunit to all cone photoreceptors
            in the mouse retina
  JOURNAL   J. Comp. Neurol. 421 (1), 80-94 (2000)
  MEDLINE   20273944
REFERENCE   2  (bases 1 to 2781)
  AUTHORS   Hirano,A.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-1999) Hirano A.A., Neuroanatomie,
            Max-Planck-Institut fuer Hirnforschung, Deutschordenstrasse 46,
            D-60528 Frankfurt a.M., GERMANY
FEATURES             Location/Qualifiers
     source          1. .2781
                     /organism="Mus musculus"
                     /plasmid="pNC-2"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /tissue_type="retina"
                     /dev_stage="adult"
     gene            380. .2275
                     /gene="CNG3"
     CDS             380. .2275
                     /gene="CNG3"
                     /function="phototransduction in cone photoreceptors"
                     /codon_start=1
                     /evidence=experimental
                     /product="cyclic nucleotide-gated channel alpha-subunit"
                     /protein_id="CAB89685.1"
                     /db_xref="GI:7688041"
                     /translation="MAKVNTQCSQPSPTQLSIKNADRDLDHVENGLGRVSRLIISIRA
                     WASRHLHDEDQTPDSFLDRFHGSELKEVSTRESNAQPNPGEQKPPDGGEGRKEEPIVV
                     DPSSNIYCRWLTAIALPVFYNWCLLVCRACFDELQSEHLTLWLVLDYSADVLYVLDML
                     VRARTGFLEQGLMVRDTKRLWKHYTKTLHFKLDILSLIPTDLAYLKLGVNYPELRFNR
                     LLKFSRLFEFFDRTETRTNYPNVFRIGNLVLYTLIIIHWNACIYFAISKFIGFGTDSW
                     VYPNTSKPEYARLSRKYIYSLYWSTLTLTTIGETPPPVKDEEYLFVVIDFLVGILIFA
                     TIVGNVGSMISNMNAPRVEFQAKIDSVKQYMQFRKVTKDLETRVIRWFDYLWANRKTV
                     DEKEVLKNLPDKLKAEIAINVHLDTLKKVRIFQDCEAGLLVELVLKLRPTVFSPGDYI
                     CKKGDIGREMYIIKEGKLAVVADDGVTQFVVLSDGSYFGEISILNIKGSKSGNRRTAN
                     IRSIGYSDLFCLSKDDLMEALTEYPDAKRALEEKGRQILMKDNLIDEDLVAARVDTRD
                     VEEKVEYLESSLDILQTRFARLLAEYSASQMKLKQRLTRLESQMNRRCCGFSPDRENS
                     EDASKTD"
BASE COUNT      672 a    707 c    743 g    659 t
ORIGIN
```

```
  Query Match           33.8%;  Score 584.8;  DB 10;  Length 2781;
  Best Local Similarity 61.4%;  Pred. No. 4.5e-126;
  Matches  975;  Conservative    0;  Mismatches  607;  Indels    6;  Gaps    2;

Qy      22 aagacaacagagtccagtcccccagccccatccaaggccaggaagttgctgcc---tgtc 78
           ||  ||  |||     |  || |||| |  |    | | |||||||  |  |||   |||
Db     623 AACCCAGGAGAACAGAAGCCACCAGACGGAGGGGAAGGCAGGAAGGAGGAGCCCATTGTG 682

Qy      79 ctggacccatctggggattactactactggtggctgaacacaatggtcttcccagtcatg 138
           |||||||  ||  |  |  |||||  | | |||||||   | ||  |   |||||||| |
Db     683 GTGGACCCCTCCAGCAACATCTACTGCCGCTGGCTGACTGCCATCGCGCTCCCAGTCTTC 742

Qy     139 tataacctcatcatcctcgtgtgcagagcctgcttccccgacttgcagcacggttatctg 198
           ||||||        | || || ||||| ||||| ||     ||  | ||    |  | |||
Db     743 TATAACTGGTGTCTACTTGTATGCAGGGCCTGTTTTGATGAGCTACAATCAGAGCACCTG 802

Qy     199 gtggcctggttggtgctggactacacgagtgacctgctatacctactagacatggtggtg 258
           |   ||| ||||  |||||||| |    ||  | || ||  |  |||||||| ||||
Db     803 ACGCTGTGGCTGGTCTTGGACTACTCTGCAGATGTCCTTTATGTTTTAGACATGCTGGTT 862

Qy     259 cgcttccacacaggattcttggaacagggcatcctggtggtggacaagggtaggatctcg 318
           ||   ||  ||||| ||| | || || ||| |  ||||    || |     ||| | | |
Db     863 CGAGCCCGGACAGGTTTCCTTGAGCAAGGCCTAATGGTCAGAGATACCAAGAGGCTGTGG 922

Qy     319 agtcgctacgttcgcacctggagtttcttcttggacctggcttccctgatgcccacagat 378
           |  |  |||       |||| |   |||    ||||| |    || || || ||||||||
Db     923 AAACATTACACAAAGACCTTGCACTTCAAGCTGGACATCCTGTCTCTCATCCCCACAGAC 982

Qy     379 gtggtctacgtgcggctgggcccgcacacacccaccctgaggctgaaccgctttctccgc 438
           |||  ||  |   | |||||  | ||   ||    ||||||  | ||||||  | ||
Db     983 CTGGCTTATTTAAAGTTGGGCGTGAACTACCCAGAACTGAGGTTCAACCGCCTCCTGAAG 1042

Qy     439 gcgccccgcctcttcgaggccttcgaccgcacagagacccgcacagcttacccaaatgcc 498
           |  || ||||| ||   ||| ||||||||||||||  | ||    |||||  || |
Db    1043 TTCTCTCGGCTCTTTGAATTCTTTGACCGCACAGAGACAAGGACCAACTACCCCAACGTG 1102

Qy     499 tttcgcattgccaagctgatgctttacatttttgtcgtcatccattggaacagctgccta 558
           ||  | || |  ||  || | |||||||    |  || |||||||  |||||   |||| |
Db    1103 TTCAGGATAGGGAACTTGGTCCTTTACACACTCATCATCATCCACTGGAATGCCTGCATC 1162

Qy     559 tactttgccctatcccggtacctgggcttcgggcgtgacgcatgggtgtacccggacccc 618
           |||||||||  | |||  ||  | | || || |||   |||  | |||||  || ||  || ||
Db    1163 TACTTTGCCATTTCCAAGTTCATTGGTTTTGGGACAGACTCCTGGGTCTATCCAAACACC 1222

Qy     619 gcgcagcctggctttgagcgcctgcggcgccagtacctctatagcttttacttctccacg 678
           |  |||| |  | ||  || ||    ||  ||||| | || ||  | ||||  |||||
Db    1223 TCCAAGCCGGAGTATGCACGACTCTCCCGGAAGTACATTTACAGTCTCTACTGGTCCACC 1282

Qy     679 ctgatactgactacagtgggcgatacaccgccgccagccagggaagaagagtacctcttc 738
           ||||  ||||| ||  | || || || || || || |  |  || |||||||| |||||
Db    1283 CTGACCCTGACCACCATTGGTGAGACCCCACCCCCCGTGAAAGATGAAGAGTATCTCTTT 1342

Qy     739 atggtgggcgacttcctgctggccgtcatgggtttcgccaccatcatgggtagcatgagc 798
```

```
Db   1343 GTGGTCATCGACTTCCTGGTGGGTATCCTGATCTTTGCCACCATAGTGGGCAACGTAGGC 1402

Qy    799 tctgtcatctacaacatgaacactgcagatgcggctttctacccagatcatgcactggtg 858
Db   1403 TCCATGATTTCCAACATGAACGCTCCCCGGGTAGAGTTCCAGGCTAAGATCGACTCCGTC 1462

Qy    859 aagaagtacatgaagctgcagcacgtcaaccgcaagctggagcggcgagttattgactgg 918
Db   1463 AAGCAGTACATGCAGTTCCGGAAGGTAACCAAGGACTTGGAGACTCGGGTTATCCGGTGG 1522

Qy    919 tatcagcacctgcagatcaacaagaagatgaccaacgaggtagccatcttacagcacttg 978
Db   1523 TTTGACTATCTGTGGGCCAACAGGAAGACGGTGGATGAAAAGGAAGTGCTCAAAAACCTC 1582

Qy    979 cctgagcggctgcgggcagaagtggctgtgtctgtgcacctgtccactctgagccgggtg 1038
Db   1583 CCGGACAAGCTGAAGGCTGAGATCGCCATCAACGTGCACCTGGACACGCTGAAGAAGGTC 1642

Qy   1039 cagatctttcagaactgtgaggccagcctgctggaggagctggtgctgaagctgcagccc 1098
Db   1643 CGAATCTTCCAGGACTGCGAGGCGGGGCTGCTGGTGGAGCTGGTGCTGAAGCTCCGCCCC 1702

Qy   1099 cagacctactcaccaggtgaatatgtatgccgcaaaggagacattggccaagagatgtac 1158
Db   1703 ACTGTGTTCAGCCCCGGGGACTACATATGCAAAAAGGGGGACATTGGAAGAGAGATGTAC 1762

Qy   1159 atcatccgagagggtcaactggccgtggtggcagatgatggtatcacacagtatgctgtg 1218
Db   1763 ATCATCAAGGAGGGCAAGCTGGCTGTGGTGGCTGATGACGGGGTCACCCAGTTTGTGGTC 1822

Qy   1219 ctcggtgcagggctctactttggggagatcagcatcatcaacatcaaagggaacatgtct 1278
Db   1823 CTCAGTGACGGCAGTTACTTTGGGGAGATTAGCATCTTAAACATTAAGGGGAGCAAGTCT 1882

Qy   1279 gggaaccgccgcacagccaacatcaagagcctaggttattcagacctattctgcctgagc 1338
Db   1883 GGGAACCGCAGGACGGCCAACATCAGGAGTATCGGCTACTCAGACCTGTTCTGCCTCTCC 1942

Qy   1339 aaggaggacctgcgggaggtgctgagcgagtatccacaagcacagaccatcatggaggag 1398
Db   1943 AAGGATGACCTAATGGAAGCCCTCACCGAGTACCCAGACGCTAAGAGGGCTCTGGAGGAA 2002

Qy   1399 aaaggacgtgagatcctgctgaaaatgaacaagttggacgtgaatgctgaggcagctgag 1458
Db   2003 AAGGGCCGTCAGATTCTGATGAAGGACAACCTAATTGATGAGGACCTAGTCGCGGCCAGG 2062

Qy   1459 atcgccctgcaggaggccacagagtcccggctacgaggcctagaccagcagctggatgat 1518
Db   2063 GTAGATACCAGGGA---CGTTGAGGAGAAGGTGGAGTACCTGGAGTCGTCCCTGGACATC 2119

Qy   1519 ctacagaccaagtttgctcgcctcctggctgagctggagtccagcgcacttaagattgct 1578
Db   2120 CTGCAGACGAGGTTTGCCCGACTCCTGGCTGAGTACAGTGCCTCCCAGATGAAGCTGAAA 2179

Qy   1579 taccgcattgaacggctggagtggcaga 1606
Db   2180 CAGCGGCTCACTCGGCTGGAGAGCCAGA 2207


RESULT  13
BTCAMPGC
LOCUS       BTCAMPGC                3166 bp    mRNA    linear   MAM 11-MAR-1993
DEFINITION  B.taurus mRNA for cAMP-gated channel.
ACCESSION   X55010
VERSION     X55010.1  GI:287743
KEYWORDS    cAMP-grated channel.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 3166)
  AUTHORS   Ludwig,J., Margalit,T., Eismann,E., Lancet,D. and Kaupp,U.B.
  TITLE     Primary structure of cAMP-gated channel from bovine olfactory
            epithelium
  JOURNAL   FEBS Lett. 270 (1-2), 24-29 (1990)
  MEDLINE   91032022
FEATURES             Location/Qualifiers
     source          1. .3166
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /cell_type="olfactory epithelium"
     CDS             418. .2409
                     /codon_start=1
                     /product="cAMP-gated channel"
                     /protein_id="CAA38754.1"
                     /db_xref="GI:287744"
                     /db_xref="SWISS-PROT:Q03041"
                     /translation="MTEKANGVKSSPANNHNHHAPPAIKASGKDDHRASSRPQSAAAD
                     DTSSELQQLAEMDAPQQRRGGFRRIARLVGVLREWAYRNFREEEPRPDSFLERFRGPE
                     LHTVTTQQGDGKGDKDGEGKGTKKKFELFVLDPAGDWYYRWLFLIALPVLYNWCLLVA
                     RACFSDLQKGYYIVWLVLDYVSDVVYIADLFIRLRTGFLEQGLLVKDTKKLRDNYIHT
                     MQFKLDVASIIPTDLIYFAVGIHNPEVRFNRLLHFARMFEFFDRTETRTSYPNIFRIS
                     NLILYILIIIHWNACIYYAISKSIGFGVDTWVYPNITDPEYGYLSREYIYCLYWSTLT
                     LTTIGETPPPVKDEEYLFVIFDFLIGVLIFATIVGNVGSMISNMNATRAEFQAKIDAV
                     KHYMQFRKVSKEMEAKVIRWFDYLWTNKKSVDEREVLKNLPAKLRAEIAINVHLSTLK
                     KVRIFQDCEAGLLVELVLKLRPQVFSPGDYICRKGDIGKEMYIIKEGKLAVVADDGVT
                     QYALLSAGSCFGEISILNIKGSKMGNRRTANIRSLGYSDLFCLSKDDLMEAVTEYPDA
                     KRVLEERGREILMKEGLLDENEVAASMEVDVQEKLEQLETNMDTLYTRFARLLAEYTG
                     AQQKLKQRITVLETKMKQNNEDDSLSDGMNSPEPPAEKP"
BASE COUNT      734 a    831 c    883 g    718 t
ORIGIN


  Query Match             33.8%;  Score 583.2;  DB 4;  Length 3166;
  Best Local Similarity   61.5%;  Pred. No. 1.1e-125;
  Matches  955;  Conservative    0;  Mismatches  593;  Indels    6;  Gaps    1;

Qy      53 ccaaggccaggaagttgctgcctgtcctggacccatctggggattactactactggtggc 112
Db     788 CCAAGAAGAAGTTTGAACTCTTTGTCTTGGACCCAGCCGGGGACTGGTACTACCGCTGGC 847

Qy     113 tgaacacaatggtcttcccagtcatgtataacctcatcatcctcgtgtgcagagcctgct 172
Db     848 TTTTTCTCATTGCCTTGCCCGTCCTCTACAACTGGTGCCTATTGGTGGCCAGAGCCTGCT 907

Qy     173 tccccgacttgcagcacggttatctggtggcctggttggtgctggactacacgagtgacc 232
Db     908 TCAGTGACCTGCAGAAAGGCTACTACATAGTGTGGCTGGTGCTGGATTACGTCTCAGATG 967

Qy     233 tgctatacctactagacatggtggtgcgcttccacacaggattcttggaacagggcatcc 292
Db     968 TGGTCTACATCGCAGACCTCTTCATCCGACTGCGCACAGGTTTCTTGGAGCAGGGGCTAC 1027

Qy     293 tggtggtggacaagggtaggatctcgagtcgctacgttcgcacctggagtttcttcttgg 352
Db    1028 TGGTGAAAGACACCAAGAAGTTGCGGGACAACTACATCCACACCATGCAGTTTAAGCTGG 1087

Qy     353 acctggcttccctgatgcccacagatgtggtctacgtgcggctgggcccgcacacaccca 412
Db    1088 ATGTGGCCTCCATCATCCCTACAGACCTGATCTATTTGCTGTGGGGATCCATAACCCTG 1147

Qy     413 ccctgaggctgaaccgctttctccgcgcgccccgcctcttcgaggccttcgaccgcacag 472
Db    1148 AGGTGCGCTTCAACCGCCTGCTACACTTTGCCCGCATGTTTGAGTTCTTTGACCGCACTG 1207

Qy     473 agacccgcacagcttacccaaatgcctttcgcattgccaagctgatgctttacatttttg 532
Db    1208 AGACACGCACCAGCTACCCCAACATCTTCCGAATAAGCAACCTGATCCTCTACATCTTGA 1267

Qy     533 tcgtcatccattggaacagctgcctatactttgccctatcccggtacctgggcttcgggc 592
Db    1268 TCATCATTCACTGGAATGCCTGCATCTACTATGCCATCTCCAAGTCCATCGGCTTTGGGG 1327

Qy     593 gtgacgcatgggtgtacccggaccccgcgcagcctggctttgagcgcctgcggcgccagt 652
```

```
Db   1328 TAGACACCTGGGTTTACCCCAACATCACTGACCCTGAGTATGGCTACCTGTCTAGGGAGT 1387

Qy    653 acctctatagcttttacttctccacgctgatactgactacagtgggcgatacaccgccgc 712
          || ||||| || |||||| || || |||| || || ||  | || || ||||| || |
Db   1388 ACATCTATTGCCTTTACTGGTCTACACTGACCCTCACCACCATTGGGGAGACACCACCCC 1447

Qy    713 cagccagggaagaagagtacctcttcatggtgggcgacttcctgctggccgtcatgggtt 772
          | |  | ||| || |||||||| ||  |  |    ||||||||| | |  ||| |    |
Db   1448 CTGTAAAGGATGAGGAGTACCTGTTTGTCATCTTTGACTTCCTGATTGGTGTCCTCATCT 1507

Qy    773 tcgccaccatcatgggtagcatgagctctgtcatctacaacatgaacactgcagatgcgg 832
          | ||||||||| |||| |   || |||  | |||| |||||||||  |  |    || |
Db   1508 TTGCCACCATCGTGGGAAATGTGGGCTCCATGATCTCCAACATGAATGCCACCCGGGCTG 1567

Qy    833 ctttctacccagatcatgcactggtgaagaagtacatgaagctgcagcacgtcaaccgca 892
          |||  |  |  |   ||     || ||  | || ||| || | |   | |||| |
Db   1568 AGTTCCAGGCCAAGATTGATGCTGTCAAACATTATATGCAGTTCCGAAAGGTCAGCAAGG 1627

Qy    893 agctggagcggcgagttattgactggtatcagcacctgcagatcaacaagaagatgacca 952
          || ||||       || |||   |||| | |  || ||  || ||| |||||||
Db   1628 AGATGGAAGCCAAGGTCATTAGGTGGTTTGACTACTTGTGGACCAATAAGAAGAGTGTAG 1687

Qy    953 acgaggtagccatcttacagcacttgcctgagcggctgcgggcagaagtggctgtgtctg 1012
          | |||  ||   || | |  || |||| |    |||  |||| || | ||  |   |
Db   1688 ATGAGCGAGAAGTCCTCAAAAACCTGCCAGCAAAGCTCAGGGCTGAGATAGCCATCAACG 1747

Qy   1013 tgcacctgtccactctgagccgggtgcagatctttcagaactgtgaggccagcctgctgg 1072
          | |||||||||||  || |     ||||  |||||||||  ||||||||||  |||||||||
Db   1748 TCCACCTGTCCACACTCAAGAAAGTGCGCATCTTTCAGGACTGTGAGGCTGGCCTGCTGG 1807

Qy   1073 aggagctggtgctgaagctgcagccccagacctactcaccaggtgaatatgtatgccgca 1132
           ||| |||||  | |||||  | ||| |||  ||      || || || ||  | |||||||
Db   1808 TGGAACTGGTATTAAAGCTCCGGCCTCAGGTCTTTAGCCCTGGGGACTACATTTGCCGCA 1867

Qy   1133 aaggagacattggccaagagatgtacatcatccgagagggtcaactggccgtggtggcag 1192
          | || || |||||  | ||||||||||| |||  |||||  || |||  | |||||| ||||||||
Db   1868 AGGGGGATATTGGGAAGGAGATGTACATAATCAAGGAGGGAAAATTGGCAGTGGTGGCTG 1927

Qy   1193 atgatggtatcacacagtatgctgtgctcggtgcagggctctactttggggagatcagca 1252
          |||| ||| |||| |||||||| |||||  || |||   | |||||| |||||||| |
Db   1928 ATGACGGTGTCACTCAGTATGCCCTGCTCTCGGCTGGGAGTTGCTTTGGAGAGATCAGTA 1987

Qy   1253 tcatcaacatcaaagggaacatgtctgggaaccgccgcacagccaacatcaagagcctag 1312
          || | || || || || | ||      || || || ||||||||||||||||  || || |
Db   1988 TCCTTAATATTAAGGGCAGCAAAATGGGCAATCGGCGCACAGCCAACATCCGCAGTCTTG 2047

Qy   1313 gttattcagacctattctgcctgagcaaggaggacctgcgggaggtgctgagcgagtatc 1372
          | || || || || |||||| ||  |||||| || ||   ||| |   |||  |||| | |
Db   2048 GCTACTCTGATCTGTTCTGCTTGTCCAAGGATGATCTTATGGAAGCTGTGACTGAGTACC 2107

Qy   1373 cacaagcacagaccatcatggaggagaaaggacgtgagatcctgctgaaaatgaacaagt 1432
          |  | ||  |||   || |||||||| ||| || ||||| ||| ||||   |  |  ||
Db   2108 CTGATGCCAAGAGGGTCTTGGAGGAGAGAGGCCGGGAGATTCTGATGAAGGAGGGCTTGT 2167

Qy   1433 tggacgtgaatgctgaggcagctgagatcgccctgcaggaggccacagagtcccggctac 1492
          |||| | |||||  | ||||||      |  | || |||  |      ||     ||||
Db   2168 TGGATGAGAATGAGGTGGCAGC------CAGCATGGAGGTAGATGTGCAGGAAAAGCTAG 2221

Qy   1493 gaggcctagaccagcagctggatgatctacagaccaagtttgctcgcctcctggctgagc 1552
          |   || ||     | ||||     |  | ||    |||||  ||||| ||||| |||
Db   2222 AACAGCTGGAGACCAACATGGACACCTTGTACACTCGTTTTGCCCGCCTGCTGGCCGAGT 2281

Qy   1553 tggagtccagcgcacttaagattgcttaccgcattgaacggctggagtggcaga 1606
          |      |   |  ||| |     | |||||   |    ||||  | |||
Db   2282 ACACGGGAGCCCAGCAGAAGCTCAAGCAGCGCATCACAGTTTTGGAAACGAAGA 2335


RESULT  14
AX179746
LOCUS       AX179746                1995 bp    DNA     linear   PAT 06-AUG-2001
DEFINITION  Sequence 34 from Patent WO0146258.
ACCESSION   AX179746
VERSION     AX179746.1  GI:15132110
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1995)
  AUTHORS   Baughn,M.R., Burford,N., Au-Young,J., Lu,D.A., Yang,J., Reddy,R.,
            Lal,P., Hillman,J.L., Azimzai,Y., Yue,H., Nguyen,D.B., Yao,M.G.,
            Gandhi,A.R., Tang,Y.T. and Khan,F.A.
  TITLE     Transporters and ion channels
  JOURNAL   Patent: WO 0146258-A 34 28-JUN-2001;
            Incyte Genomics, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .1995
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /note="Incyte ID No: 7472002CB1"
BASE COUNT      511 a    511 c    532 g    441 t
ORIGIN


  Query Match             33.5%;  Score 578.6;  DB 6;  Length 1995;
  Best Local Similarity   60.7%;  Pred. No. 1.2e-124;
  Matches  966;  Conservative    0;  Mismatches  619;  Indels    6;  Gaps    1;

Qy     70 ctgcctgtcctggacccatctggggattactactactggtggctgaacacaatggtcttc 129
          ||   |||| |||||||| |||||||| |||||||| |||||       || | | |
Db    388 CTATTTGTCTTGGACCCAGCTGGGGATTGGTACTACTGCTGGCTATTTGTCATTGCCATG 447

Qy    130 ccagtcatgtataacctcatcatcctcgtgtgcagagcctgcttccccgacttgcagcac 189
          || ||| | || |||     | | || |||  |||||||||||||   ||| | ||| |
Db    448 CCCGTCCTTTACAACTGGTGCCTGCTGGTGGCCAGAGCCTGCTTCAGTGACCTACAGAAA 507

Qy    190 ggttatctggtggcctggttggtgctggactacacgagtgacctgctatacctactagac 249
          || ||     |||  ||| ||||||||||| ||       ||  || | ||| |    |||
Db    508 GGCTACTACCTGGTGTGGCTGGTGCTGGATTATGTCTCAGATGTGGTCTACATTGCGGAC 567

Qy    250 atggtggtgcgcttccacacaggattcttggaacagggcatcctggtggtggacaagggt 309
          |  |  | || || | |||||| ||| |||| ||||  | |||||    || |
Db    568 CTCTTCATCCGATTGCGCACAGGTTTCCTGGAGCAGGGGCTGCTGGTCAAAGATACCAAG 627

Qy    310 aggatctcgagtcgctacgttcgcacctggagtttcttcttggacctggcttccctgatg 369
          |   |      |||| | | |||| |  |||    |||| ||||||| | ||
Db    628 AAACTGCGAGACAACTACATCCACACCCTGCAGTTCAAGCTGGATGTGGCTTCCATCATC 687

Qy    370 cccacagatgtggtctacgtgcggctgggcccgcacacacccaccctgaggctgaaccgc 429
          ||||| ||  || |||| |     ||| |   |||| ||     || |  | ||||||
Db    688 CCCACTGACCTGATCTATTTGCTGTGGACATCCACAGCCCTGAGGTGCGCTTCAACCGC 747

Qy    430 tttctccgcgcgccccgcctcttcgaggccttcgaccgcacagagacccgcacagcttac 489
          | || | |     ||||| | || |||  ||| |||||  |||||||| |||||    |||
Db    748 CTGCTGCACTTTGCCCGCATGTTTGAGTTCTTTGACCGGACAGAGACACGCACCAACTAC 807

Qy    490 ccaaatgcctttcgcattgccaagctgatgctttacatttttgtcgtcatccattggaac 549
          || ||   ||| |||||   ||| ||  | || ||||| || ||| ||||||| |||||
Db    808 CCTAACATCTTCCGCATCAGCAACCTTGTCCTCTACATCTTGGTCATCATCCACTGGAAT 867

Qy    550 agctgcctatactttgccctatcccggtacctgggcttcgggcgtgacgcatgggtgtac 609
           |||| | || | |||| | |||   | | | ||||| |||   ||| | ||||| |||
Db    868 GCCTGCATCTATTATGCCATCTCCAAATCCATAGGCTTTGGGGTCGACACCTGGGTTTAC 927

Qy    610 ccggaccccgcgcagcctggctttgagcgcctgcggcgccagtacctctatagcttttac 669
          ||  ||  | |  | |||   | ||    ||||     |  | ||| ||||| || |||||
Db    928 CCAAACATCACTGACCCTGAGTATGGCTACCTGGCTAGGGAATACATCTATTGCCTTTAC 987

Qy    670 ttctccacgctgatactgactacagtgggcgatacaccgccgccagccagggaagaagag 729
          |  ||||| ||||  || |||||  | || || ||||| || || |  | ||| || |||
Db    988 TGGTCCACACTGACTCTCACTACCATTGGGGAGACACCACCCCCTGTAAAGGATGAGGAG 1047

Qy    730 tacctcttcatggtgggcgacttcctgctggccgtcatgggtttcgccaccatcatgggt 789
          ||||| ||  |  |    ||||||||| | | |||| |    || |||||||| ||||
Db   1048 TACCTATTTGTCATCTTTGACTTCCTGATTGGCGTCCTCATCTTTGCCACCATCGTGGGA 1107
```

```
Qy    790 agcatgagctctgtcatctacaacatgaacactgcagatgcggctttctacccagatcat 849
Db   1108 AATGTGGGCTCCATGATCTCCAACATGAATGCCACCCGGGCAGAGTTCCAGGCTAAGATC 1167

Qy    850 gcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggagcggcgagtt 909
Db   1168 GATGCCGTGAAACACTACATGCAGTTCCGAAAGGTCAGCAAGGGGATGGAAGCCAAGGTC 1227

Qy    910 attgactggtatcagcacctgcagatcaacaagaagatgaccaacgaggtagccatctta 969
Db   1228 ATTAGGTGGTTTGACTACTTGTGGACCAATAAGAAGACAGTGGATGAGCGAGAAATTCTC 1287

Qy    970 cagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctgtccactctg 1029
Db   1288 AAGAATCTGCCAGCCAAGCTCAGGGCTGAGATAGCCATCAATGTCCACTTGTCCACACTC 1347

Qy   1030 agccgggtgcagatctttcagaactgtgaggccagcctgctggaggagctggtgctgaag 1089
Db   1348 AAGAAAGTGCGCATCTTCCATGATTGTGAGGCTGGCCTGCTGGTAGAGCTGGTACTGAAA 1407

Qy   1090 ctgcagccccagacctactcaccaggtgaatatgtatgccgcaaaggagacattggccaa 1149
Db   1408 CTCCGTCCTCAGGTCTTCAGTCCTGGGGATTACATTTGCCGCAAAGGGGACATCGGCAAG 1467

Qy   1150 gagatgtacatcatccgagagggtcaactggccgtggtggcagatgatggtatcacacag 1209
Db   1468 GAGATGTACATCATTAAGGAGGGCAAACTGGCAGTGGTGGCTGATGATGGTGTGACTCAG 1527

Qy   1210 tatgctgtgctcggtgcagggctctactttggggagatcagcatcatcaacatcaaaggg 1269
Db   1528 TATGCTCTGCTGTCGGCTGGAAGCTGCTTTGGCGAGATCAGTATCCTTAACATTAAGGGC 1587

Qy   1270 aacatgtctgggaaccgccgcacagccaacatcaagagcctaggttattcagacctattc 1329
Db   1588 AGTAAAATGGGCAATCGACGCACAGCTAATATCCGCAGCCTGGGCTACTCAGATCTCTTC 1647

Qy   1330 tgcctgagcaaggaggacctgcgggaggtgctgagcgagtatccacaagcacagaccatc 1389
Db   1648 TGCTTGTCCAAGGATGATCTTATGGAAGCTGTGACTGAGTACCCTGATGCCAAGAAAGTC 1707

Qy   1390 atggaggagaaaggacgtgagatcctgctgaaaatgaacaagttggacgtgaatgctgag 1449
Db   1708 CTAGAAGAGAGGGGTCGGGAGATCCTCATGAAGGAGGGACTGCTGGATGAGAACGAAGTG 1767

Qy   1450 gcagctgagatcgccctgcaggaggccacagagtcccggctacgaggcctagaccagcag 1509
Db   1768 GCA------ACCAGCATGGAGGTCGACGTGCAGGAGAAGCTAGGGCAGCTGGAGACCAAC 1821

Qy   1510 ctggatgatctacagaccaagtttgctcgcctcctggctgagctggagtccagcgcactt 1569
Db   1822 ATGGAAACCTTGTACACTCGCTTTGGCCGCCTGCTGGCTGAGTACACGGGGGCCCAGCAG 1881

Qy   1570 aagattgcttaccgcattgaacggctggagtggcagactcgagagtggccaatgcccgag 1629
Db   1882 AAGCTCAAGCAGCGCATCACAGTTCTGGAAACCAAGATGAAACAGAACAATGAAGATGAC 1941

Qy   1630 gacctggctgaggctgatgacgagggtgagc 1660
Db   1942 TACCTGTCTGATGGGATGAACAGCCCTGAGC 1972
```

```
RESULT  15
OCRACNG
LOCUS       OCRACNG                 2669 bp    mRNA    linear   MAM 08-DEC-1993
DEFINITION  O.cuniculus rACNG mRNA for aorta CNG channel.
ACCESSION   X59668
VERSION     X59668.1  GI:433959
KEYWORDS    CNG channel.
SOURCE      rabbit.
  ORGANISM  Oryctolagus cuniculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
REFERENCE   1  (bases 1 to 2669)
  AUTHORS   Biel,M., Altenhofen,W., Hullin,R., Ludwig,J., Freichel,M.,
            Flockerzi,V., Dascal,N., Kaupp,U.B. and Hofmann,F.
  TITLE     Primary structure and functional expression of a cyclic
            nucleotide-gated channel from rabbit aorta
  JOURNAL   FEBS Lett. 329 (1-2), 134-138 (1993)
  MEDLINE   93359035
REFERENCE   2  (bases 1 to 2669)
  AUTHORS   Hofmann,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-1993) F. Hofmann, Institut f. Pharmakologie &
            Toxikologie, Technische Universitaet Muenchen, Biedersteiner Str.
            29, 80802 Muenchen, FRG
FEATURES             Location/Qualifiers
     source          1. .2669
                     /organism="Oryctolagus cuniculus"
                     /db_xref="taxon:9986"
                     /tissue_type="aorta"
                     /clone_lib="plasmid vector pSport (BRL)"
                     /dev_stage="adult"
     gene            153. .2351
                     /gene="rACNG"
     CDS             153. .2351
                     /gene="rACNG"
                     /codon_start=1
                     /product="aorta CNG channel (rACNG)"
                     /protein_id="CAA42201.1"
                     /db_xref="GI:433960"
                     /db_xref="SWISS-PROT:Q28718"
                     /translation="MSSWRSCARAPLSGSAWRRSAATRRSRRCLKTKRKRWSSGKGTP
                     MQSTQCETRRRAQTPCESTGHTWRMTEKSNGVKSSPANNHNNHVPATIKANGKDESRT
                     RSRPQSAADDDTSSELQRLAEMDAPQQRRGGFRRIVRLVGVIRQWANRNFREEEARPD
                     SFLERFRGPELQTVTTQQGDGKGDKDGDGKGTKKKFELFVLDPAGDWYYRWLFVIAMP
                     VLYNWCLLVARACFSDLQRGYFLVWLVLDYFSDVVYIADLFIRLRTGFLEQGLLVKDP
                     KKLRDNYIHTLQFKLDVASIIPTDLIYFAVGIHNPELRFNRLLHFARMFEFFDRTETR
                     TSYPNIFRISNLVLYILVIIHWNACIYYAISKSIGFGVDTWVYPNITDPEYGYLAREY
                     IYCLYWSTLTLTTIGETPPPVKDEEYLFVIFDFLIGVLIFATIVGNVGSMISNMNATR
                     AEFQAKIDAVKHYMQFRKVSKEMEAKVIKWFDYLWTNKKTVDEREVLKNLPAKLRAEI
                     AINVHLSTLKKVRIFQDCEAGLLVELVLKLRPQVFSPGDYICRKGDIGKEMYIIKEGK
                     LAVVADDGVTQYALLSAGSCFGEISILNIKGSKMGNRRTANIRSLGYSDLFCLSKDDL
                     MEAVTEYPDAKKVLEERGREILMKEGLLDENEVAASMEVDVQEKLKQLETNMETLYTR
                     FGRLLAEYTGAQQKLKQRITVLEVKMKQNTEDDYLSDGMNSPEPAAAEQP"
BASE COUNT      682 a    639 c    742 g    606 t
ORIGIN
```

```
  Query Match             33.1%;  Score 572;  DB 4;  Length 2669;
  Best Local Similarity   60.4%;  Pred. No. 4.4e-123;
  Matches  963;  Conservative    0;  Mismatches  625;  Indels    6;  Gaps    1;

Qy     70 ctgcctgtcctggacccatctggggattactactactggtggctgaacacaatggtcttc 129
Db    744 CTATTTGTCTTGGACCCAGCTGGGGACTGGTACTACCGCTGGCTGTTCGTCATTGCCATG 803

Qy    130 ccagtcatgtataacctcatcatcctcgtgtgcagagcctgcttccccgacttgcagcac 189
Db    804 CCTGTGCTATACAACTGGTGTCTGCTGGTGGCCAGAGCCTGTTTCAGTGACTTGCAGAGA 863

Qy    190 ggttatctggtggcctggttggtgctggactacacgagtgacctgctatacctactagac 249
Db    864 GGCTACTTCCTGGTGTGGCTGGTGCTGGACTACTTCTCAGATGTGGTCTACATTGCGGAT 923

Qy    250 atggtggtgcgcttccacacaggattcttggaacagggcatcctggtggtggacaagggt 309
Db    924 CTTTTCATCCGATTGCGCACAGGTTTCTTGGAGCAGGGGCTGCTAGTGAAAGATCCCAAG 983

Qy    310 aggatctcgagtcgctacgttcgcacctggagtttcttcttggacctggcttccctgatg 369
Db    984 AAATTGCGAGACAACTACATCCACACCCTACAGTTCAAGCTAGATGTGGCTTCCATCATT 1043

Qy    370 cccacagatgtggtctacgtgcggctgggcccgcacacacccaccctgaggctgaaccgc 429
Db   1044 CCCACAGACCTGATCTATTTTGCTGTGGGAATCCACAACCCCGAGCTGCGCTTCAATCGT 1103
```

```
Qy    430 tttctccgcgcgccccgcctcttcgaggccttcgaccgcacagagacccgcacagcttac 489
          | || |      ||||| | || ||| ||| || ||||| || || || ||   ||||
Db   1104 CTGCTACATTTTGCCCGCATGTTTGAGTTCTTTGATCGCACTGAAACACGTACCAGTTAC 1163

Qy    490 ccaaatgcctttcgcattgccaagctgatgctttacatttttgtcgtcatccattggaac 549
          || ||   ||| ||||   ||| ||  | || |||||||| ||| ||||||| |||||
Db   1164 CCCAACATCTTCCGCATCAGCAACCTAGTCCTCTACATTTTGGTCATCATCCACTGGAAT 1223

Qy    550 agctgcctatactttgccctatcccggtacctgggcttcgggcgtgacgcatgggtgtac 609
            |||  | |||| |||  | |||  || | | ||||| |||  |||  | ||||| |||
Db   1224 GCCTGTATCTACTATGCAATCTCCAAGTCCATTGGCTTTGGGGTTGATACCTGGGTTTAC 1283

Qy    610 ccggaccccgcgcagcctggctttgagcgcctgcggcgccagtacctctatagcttttac 669
          ||  |   | |  | ||||  | ||    |||     |  ||||  ||||| |  |||||
Db   1284 CCCAATATCACTGACCCTGAATATGGATACCTAGCTAGGGAGTATATCTATTGTCTTTAC 1343

Qy    670 ttctccacgctgatactgactacagtgggcgatacaccgccgccagccagggaagaagag 729
          |  ||||| |||| || || ||  | || || ||||| || || |  |  || || |||
Db   1344 TGGTCCACACTGACCCTCACCACCATTGGGGAGACACCACCCCCTGTAAAAGATGAGGAG 1403

Qy    730 taccttcatggtgggcgacttcctgctggccgtcatgggtttcgccaccatcatgggt 789
          ||||| ||  |  |    |||||||||| | |  ||| |    || |||||||||| ||||
Db   1404 TACCTATTTGTTATCTTTGACTTCCTGATTGGTGTCCTCATCTTTGCCACCATCGTGGGA 1463

Qy    790 agcatgagctctgtcatctacaacatgaacactgcagatgcggctttctacccagatcat 849
          |   || ||||| | || | ||| ||||  |  |    || |  ||| |  |  |   |
Db   1464 AATGTGGGCTCTATGATTTCCAATATGAATGCCACCAGGGCAGAATTCCAGGCCAAAATT 1523

Qy    850 gcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggagcggcgagtt 909
          |    || ||| | |||||| || | |   | |||| |    |  ||||      ||
Db   1524 GATGCTGTCAAGCACTACATGCAGTTCCGCAAAGTCAGCAAGGAAATGGAAGCCAAGGTC 1583

Qy    910 attgactggtatcagcacctgcagatcaacaagaagatgaccaacgaggtagccatctta 969
          ||| | |||| | |  |||||  || ||| |||||||      | |||   |   || |
Db   1584 ATTAAGTGGTTTGACTACCTGTGGACCAATAAGAAGACTGTAGATGAGCGGGAAGTCCTC 1643

Qy    970 cagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctgtccactctg 1029
          || || | || |    |||  |||| ||  | ||| |   ||| ||||||||||| ||
Db   1644 AAGAACCTACCAGCCAAGCTCAGGGCTGAGATAGCTATAAATGTTCACCTGTCCACACTC 1703

Qy   1030 agccgggtgcagatctttcagaactgtgaggccagcctgctggaggagctggtgctgaag 1089
          |    || |  |||| ||| |||||||||  |||||||| |||||||||||||||
Db   1704 AAGAAAGTACGCATCTTCCAGGACTGTGAGGCTGGCCTGCTGGTGGAGCTGGTGCTGAAG 1763

Qy   1090 ctgcagccccagacctactcaccaggtgaatatgtatgccgcaaaggagacattggccaa 1149
          || |  || |||  || |   || || || ||  | |||||||| || |||||||| |
Db   1764 CTCCGTCCTCAGGTCTTCAGCCCGGGGGATTACATTTGCCGCAAGGGGGACATTGGCAAG 1823

Qy   1150 gagatgtacatcatccgagagggtcaactggccgtggtggcagatgatggtatcacacag 1209
          |||||||||| |||   || ||  |  |||| |||||||| |||||||| | || |||
Db   1824 GAGATGTACATAATCAAGGAAGGCAAGTTGGCAGTGGTGGCTGATGATGGTGTGACTCAG 1883

Qy   1210 tatgctgtgctcggtgcagggctctactttggggagatcagcatcatcaacatcaaaggg 1269
          |||||  |||||   || |||   | |||||| |||| |||||| | |||| |||||
Db   1884 TATGCCCTGCTCTCGGCTGGGAGTTGCTTTGGCGAGATTAGCATCCTTAACATTAAAGGC 1943

Qy   1270 aacatgtctgggaaccgccgcacagccaacatcaagagcctaggttattcagacctattc 1329
          | ||     || || || |||||||||||| ||   ||| || || || ||||| || |||
Db   1944 AGCAAAATGGGCAATCGACGCACAGCCAATATTCGGAGTCTGGGCTACTCAGATCTCTTC 2003

Qy   1330 tgcctgagcaaggaggacctgcgggaggtgctgagcgagtatccacaagcacagaccatc 1389
          ||| ||  |||||| |||||   ||| |   |||  |||||||  | ||  |||   ||
Db   2004 TGCTTGTCCAAGGATGACCTCATGGAAGCTGTGACTGAGTATCCTGACGCCAAGAAGGTC 2063

Qy   1390 atggaggagaaaggacgtgagatcctgctgaaaatgaacaagttggacgtgaatgctgag 1449
          ||||||||  ||  | |||||| |  ||||   |     | |||| | ||||  | |
Db   2064 TTGGAGGAGAGGGGTAGGGAGATCTTAATGAAGGAGGGGCTGCTGGATGAGAATGAGGTG 2123

Qy   1450 gcagctgagatcgccctgcaggaggccacagagtcccggctacgaggcctagaccagcag 1509
          |||||      |  | || |||  |      |      ||||      || ||     |
Db   2124 GCAGC------CAGCATGGAGGTAGATGTGCAAGAGAAGCTAAAGCAGCTGGAGACCAAC 2177

Qy   1510 ctggatgatctacagaccaagtttgctcgcctcctggctgagctggagtccagcgcactt 1569
          ||||     |  | ||    ||||  ||||| ||||| |||     |     |   |
Db   2178 ATGGAGACCTTGTACACTCGCTTTGGACGCCTGCTGGCAGAGTACACGGGGGCCCAGCAG 2237

Qy   1570 aagattgcttaccgcattgaacggctggagtggcagactcgagagtggccaatgcccgag 1629
          ||  |     | |||||       ||||||    |||    | ||    |   |   ||
Db   2238 AAACTCAAGCAGCGCATCACGGTTCTGGAGGTCAAGATGAAACAGAACACTGAGGATGAT 2297

Qy   1630 gacctggctgaggctgatgacgagggtgagcctg 1663
          |  || | || |      ||     ||||||  |
Db   2298 TATTTGTCAGATGGAATGAACAGTCCTGAGCCAG 2331
```

Search completed: June 20, 2002, 14:35:16
Job time: 19579 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2002, 14:38:56 ; Search time 668.44 Seconds
(without alignments)
4438.436 Million cell updates/sec

Title: US-09-927-267-3
Perfect score: 1728
Sequence: 1 atgagccaggacaccaaagt.....gacccccaggtccagagtga 1728

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1726.4	99.9	2551	24	ABA04603	MOL10b coding sequ
	2	1687.4	97.7	1835	24	ABA04602	MOL10a coding sequ
	3	578.6	33.5	1995	22	AAD09558	Human transporter
	4	528.6	30.6	995	22	AAH98838	Rat EST-derived co
	5	405.2	23.4	2909	23	ABL13353	Drosophila melanog
	6	279.2	16.2	2109	23	ABL18085	Drosophila melanog
	7	222.2	12.9	5133	23	ABL14931	Drosophila melanog
	8	191.2	11.1	8585	23	ABL13352	Drosophila melanog
	9	169.2	9.8	1125	23	AAS91258	DNA encoding novel
	10	163.6	9.5	1281	23	ABL06311	Drosophila melanog
c	11	141	8.2	17150	23	ABL18084	Drosophila melanog
	12	137	7.9	3275	23	ABL19457	Drosophila melanog
	13	124.8	7.2	3811	23	AAS91658	DNA encoding novel
	14	118.4	6.9	1146	23	AAS79945	DNA encoding novel
	15	118.4	6.9	2421	23	AAS77668	DNA encoding novel
c	16	91.8	5.3	5692	23	ABL19456	Drosophila melanog
	17	89.2	5.2	4751	20	AAZ09496	Human heart tissue
	18	89.2	5.2	5065	22	AAH48730	Human HCN4 cDNA.
	19	89.2	5.2	5499	22	ABA09197	Human cation chann
	20	79	4.6	2757	24	AAD25729	Human cyclic nucle
	21	78.4	4.5	1307	20	AAX25488	Human cyclic nucle
c	22	76.2	4.4	12770	23	ABL14930	Drosophila melanog
	23	72.4	4.2	1820	20	AAZ09488	Bovine retina Ih i
	24	70.8	4.1	1342	20	AAZ09482	Human Ih ion chann
	25	70.2	4.1	1507	20	AAX84446	Mouse brain CNG-3
	26	69.2	4.0	3372	22	AAH48729	Human HCN2 cDNA.
	27	65.8	3.8	1790	20	AAX84445	Human brain CNG-2
	28	65	3.8	2886	20	AAZ09491	Human Ih ion chann
	29	62.6	3.6	3431	22	AAH98302	Human EST-derived
c	30	62.4	3.6	5645	23	ABL06310	Drosophila melanog
	31	60.4	3.5	2325	21	AAC66779	Human hyperpolaris
	32	59.8	3.5	3479	22	AAH21451	Human HERG1 DNA.
	33	59.8	3.5	3480	21	AAA07601	Long QT syndrome a
	34	59.8	3.5	3950	21	AAA07602	Long QT syndrome a
	35	59.8	3.5	4070	22	AAI66257	Human potassium ch
	36	59.8	3.5	4070	22	AAF61965	Human eag-related
	37	56	3.2	1584	20	AAX84444	Mouse brain CNG-2
	38	56	3.2	3102	22	AAH48731	Murine HCN2 cDNA.
	39	55.2	3.2	3438	21	AAA62050	Hydrophobic domain
	40	53.6	3.1	3355	21	AAZ50452	Monkey potassium c
	41	52.2	3.0	794	20	AAZ09494	Rat olfactory Ih i
	42	52	3.0	3249	21	AAZ50119	Human Elk voltage
	43	52	3.0	3252	21	AAA14893	DNA encoding a her
	44	52	3.0	3323	20	AAX84910	Human brain specif
	45	52	3.0	3829	21	AAZ87712	Human ESK1 (hESK1)

ALIGNMENTS

RESULT 1
ABA04603
ID ABA04603 standard; cDNA; 2551 BP.
XX
AC ABA04603;
XX
DT 21-FEB-2002 (first entry)
XX
DE MOL10b coding sequence.
XX
KW MOL; G-coupled protein-receptor; cardiomyopathy; atherosclerosis;
KW cell signal processing; metabolic disorder; diabetes; cancer; human;
KW neurodegenerative disorder; immune disorder; cardiac disorder;
KW lung disease; autoimmune disease; developmental disorder; antidiabetic;
KW Cytostatic; Neuroprotective; Antiatherosclerotic; Immunosuppressive;
KW Gene therapy; Vaccine; antiinflammatory; MOL10b; chromosome 11; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 779..2506
FT /*tag= a
FT /product= "MOL 10b"
XX
PN WO200181578-A2.
XX
PD 01-NOV-2001.
XX
PF 26-APR-2001; 2001WO-US13578.
XX
PR 26-APR-2000; 2000US-200158P.
PR 28-APR-2000; 2000US-200613P.

```
PR   28-APR-2000; 2000US-200780P.
PR   01-MAY-2000; 2000US-201006P.
PR   01-MAY-2000; 2000US-201007P.
PR   01-MAY-2000; 2000US-201236P.
PR   01-MAY-2000; 2000US-201238P.
PR   02-MAY-2000; 2000US-201186P.
PR   03-MAY-2000; 2000US-201474P.
PR   03-MAY-2000; 2000US-201508P.
PR   25-JUL-2000; 2000US-220591P.
PR   15-SEP-2000; 2000US-232678P.
PR   22-JAN-2001; 2001US-263217P.
PR   30-JAN-2001; 2001US-265160P.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Vernet CAM,  Fernandes ER,  Gerlach V,  Shimkets RA,  Malyankar UM;
PI   Boldog FL,  Zerhusen BD,  Spytek KA,  Majumder K,  Tchernev VT;
PI   Padigaru M,  Patturajan M,  Burgess CE,  Gangolli EA,  Smithson G;
PI   Rastelli L,  Macdougall JR,  Taupier RJ,  Grosse WM,  Szekeres ES;
PI   Alsobrook JP;
XX
DR   WPI; 2002-049278/06.
DR   P-PSDB; AAM47673.
XX
PT   Novel G-protein coupled receptor-related polypeptides and
PT   polynucleotides for diagnosing, preventing and treating cardiomyopathy,
PT   atherosclerosis, disorders related to cell signal processing and for
PT   identifying modulators  -
XX
PS   Claim 8; Page 78; 227pp; English.
XX
CC   The present invention relates to novel G-coupled protein-receptor related
CC   proteins and coding sequences (MOLX, where X is a number from 1 to 10,
CC   ABA04589-ABA04603 and AAM47659-AAM47673). MOLX proteins and coding
CC   sequences are useful for treating or preventing a MOLX-associated
CC   disorder, such as cardiomyopathy, atherosclerosis, disorders related to
CC   cell signal processing and metabolic pathway modulation, diabetes and
CC   cancer. Additionally, MOLX proteins and coding sequences are useful for
CC   preventing and treating a variety of disorders including metabolic
CC   disorders, nutritional oedema, chronic and hereditary pancreatitis,
CC   obesity, infectious disease, anorexia, neurodegenerative disorders,
CC   Alzheimer's disease, Parkinson's disease, stroke, immune disorders,
CC   haematopoietic disorders and various dyslipidaemias, metabolic syndrome X
CC   and wasting disorders associated with chronic diseases and cancers,
CC   cardiac disorders, hypertension, hypercalcaemia, cirrhosis, angiogenesis
CC   and wound healing, trauma, glomerulonephritis, hyper and hypothyroidism,
CC   multiple sclerosis, lung diseases including asthma, Crohn's disease,
CC   scleroderma, autoimmune diseases, developmental disorders and neural tube
CC   defects. The present sequence is the coding sequence for human MOL10b,
CC   which maps to chromosome 11. MOL10b protein is a cyclic nucleotide-gated
CC   olfactory channel-like protein.
XX
SQ   Sequence 2551 BP; 646 A; 668 C; 707 G; 530 T; 0 other;

  Query Match             99.9%;  Score 1726.4;  DB 24;  Length 2551;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1727;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 atgagccaggacaccaaagtgaagacaacagagtccagtcccccagccccatccaaggcc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    779 atgagccaggacaccaaagtgaagacaacagagtccagtcccccagccccatccaaggcc 838

Qy     61 aggaagttgctgcctgtcctggacccatctggggattactactactggtggctgaacaca 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    839 aggaagttgctgcctgtcctggacccatctggggattactactactggtggctgaacaca 898

Qy    121 atggtcttcccagtcatgtataacctcatcatcctcgtgtgcagagcctgcttccccgac 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    899 atggtcttcccagtcatgtataacctcatcatcctcgtgtgcagagcctgcttccccgac 958

Qy    181 ttgcagcacggttatctggtggcctggttggtgctggactacacgagtgacctgctatac 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    959 ttgcagcacggttatctggtggcctggttggtgctggactacacgagtgacctgctatac 1018

Qy    241 ctactagacatggtggtgcgcttccacacaggattcttggaacagggcatcctggtggtg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1019 ctactagacatggtggtgcgcttccacacaggattcttggaacagggcatcctggtggtg 1078

Qy    301 gacaagggtaggatctcgagtcgctacgttcgcacctggagtttcttcttggacctggct 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1079 gacaagggtaggatctcgagtcgctacgttcgcacctggagtttcttcttggacctggct 1138

Qy    361 tccctgatgcccacagatgtggtctacgtgcggctgggcccgcacacacccacccctgagg 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1139 tccctgatgcccacagatgtggtctacgtgcggctgggcccgcacacacccacccctgagg 1198

Qy    421 ctgaaccgctttctccgcgcgccccgcctcttcgaggccttcgaccgcacagagacccgc 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1199 ctgaaccgctttctccgcgcgccccgcctcttcgaggccttcgaccgcgcagagacccgc 1258

Qy    481 acagcttacccaaatgcctttcgcattgccaagctgatgctttacatttttgtcgtcatc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1259 acagcttacccaaatgcctttcgcattgccaagctgatgctttacatttttgtcgtcatc 1318

Qy    541 cattggaacagctgcctatactttgccctatcccggtacctgggcttcgggcgtgacgca 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1319 cattggaacagctgcctatactttgccctatcccggtacctgggcttcgggcgtgacgca 1378

Qy    601 tgggtgtacccggaccccgcgcagcctggctttgagcgcctgcggcgccagtacctctat 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1379 tgggtgtacccggaccccgcgcagcctggctttgagcgcctgcggcgccagtacctctat 1438

Qy    661 agcttttacttctccacgctgatactgactacagtgggcgatacaccgccgccagccagg 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1439 agcttttacttctccacgctgatactgactacagtgggcgatacaccgccgccagccagg 1498

Qy    721 gaagaagagtacctcttcatggtgggcgacttcctgctggccgtcatgggtttcgccacc 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1499 gaagaagagtacctcttcatggtgggcgacttcctgctggccgtcatgggtttcgccacc 1558

Qy    781 atcatgggtagcatgagctctgtcatctacaacatgaacactgcagatgcggctttctac 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1559 atcatgggtagcatgagctctgtcatctacaacatgaacactgcagatgcggctttctac 1618

Qy    841 ccagatcatgcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggag 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1619 ccagatcatgcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggag 1678

Qy    901 cggcgagttattgactggtatcagcacctgcagatcaacaagaagatgaccaacgaggta 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1679 cggcgagttattgactggtatcagcacctgcagatcaacaagaagatgaccaacgaggta 1738

Qy    961 gccatcttacagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctg 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1739 gccatcttacagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctg 1798

Qy   1021 tccactctgagccgggtgcagatctttcagaactgtgaggccagcctgctggaggagctg 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1799 tccactctgagccgggtgcagatctttcagaactgtgaggccagcctgctggaggagctg 1858

Qy   1081 gtgctgaagctgcagccccagacctactcaccaggtgaatatgtatgccgcaaaggagac 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1859 gtgctgaagctgcagccccagacctactcaccaggtgaatatgtatgccgcaaaggagac 1918

Qy   1141 attggccaagagatgtacatcatccgagagggtcaactggccgtggtggcagatgatggt 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1919 attggccaagagatgtacatcatccgagagggtcaactggccgtggtggcagatgatggt 1978

Qy   1201 atcacacagtatgctgtgctcggtgcagggctctactttggggagatcagcatcatcaac 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1979 atcacacagtatgctgtgctcggtgcagggctctactttggggagatcagcatcatcaac 2038

Qy   1261 atcaaagggaacatgtctgggaaccgccgcacagccaacatcaagagcctaggttattca 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2039 atcaaagggaacatgtctgggaaccgccgcacagccaacatcaagagcctaggttattca 2098
```

```
Qy   1321 gacctattctgcctgagcaaggaggacctgcgggaggtgctgagcgagtatccacaagca 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2099 gacctattctgcctgagcaaggaggacctgcgggaggtgctgagcgagtatccacaagca 2158

Qy   1381 cagaccatcatggaggagaaaggacgtgagatcctgctgaaaatgaacaagttggacgtg 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2159 cagaccatcatggaggagaaaggacgtgagatcctgctgaaaatgaacaagttggacgtg 2218

Qy   1441 aatgctgaggcagctgagatcgccctgcaggaggccacagagtcccggctacgaggccta 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2219 aatgctgaggcagctgagatcgccctgcaggaggccacagagtcccggctacgaggccta 2278

Qy   1501 gaccagcagctggatgatctacagaccaagtttgctcgcctcctggctgagctggagtcc 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2279 gaccagcagctggatgatctacagaccaagtttgctcgcctcctggctgagctggagtcc 2338

Qy   1561 agcgcacttaagattgcttaccgcattgaacggctggagtggcagactcgagagtggcca 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2339 agcgcacttaagattgcttaccgcattgaacggctggagtggcagactcgagagtggcca 2398

Qy   1621 atgcccgaggacctggctgaggctgatgacgagggtgagcctgaggagggaacttccaaa 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2399 atgcccgaggacctggctgaggctgatgacgagggtgagcctgaggagggaacttccaaa 2458

Qy   1681 gatgaagagggcagggccagccaggagggacccccaggtccagagtga 1728
          ||||||||||||||||||||||||||||||||||||||||||||||
Db   2459 gatgaagagggcagggccagccaggagggacccccaggtccagagtga 2506
```

```
RESULT   2
ABA04602
ID   ABA04602 standard; cDNA; 1835 BP.
XX
AC   ABA04602;
XX
DT   21-FEB-2002  (first entry)
XX
DE   MOL10a coding sequence.
XX
KW   MOL; G-coupled protein-receptor; cardiomyopathy; atherosclerosis;
KW   cell signal processing; metabolic disorder; diabetes; cancer;
KW   neurodegenerative disorder; immune disorder; cardiac disorder;
KW   lung disease; autoimmune disease; developmental disorder; antidiabetic;
KW   Cytostatic; Neuroprotective; Antiatherosclerotic; Immunosuppressive;
KW   Gene therapy; Vaccine; antiinflammatory; MOL10a; ss.
XX
OS   Unidentified.
XX
FH   Key             Location/Qualifiers
FT   CDS             54..1790
FT                   /*tag= a
FT                   /product= "MOL 10a"
XX
PN   WO200181578-A2.
XX
PD   01-NOV-2001.
XX
PF   26-APR-2001; 2001WO-US13578.
XX
PR   26-APR-2000; 2000US-200158P.
PR   28-APR-2000; 2000US-200613P.
PR   28-APR-2000; 2000US-200780P.
PR   01-MAY-2000; 2000US-201006P.
PR   01-MAY-2000; 2000US-201007P.
PR   01-MAY-2000; 2000US-201236P.
PR   01-MAY-2000; 2000US-201238P.
PR   02-MAY-2000; 2000US-201186P.
PR   03-MAY-2000; 2000US-201474P.
PR   03-MAY-2000; 2000US-201508P.
PR   25-JUL-2000; 2000US-220591P.
PR   15-SEP-2000; 2000US-232678P.
PR   22-JAN-2001; 2001US-263217P.
PR   30-JAN-2001; 2001US-265160P.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Vernet CAM,  Fernandes ER,  Gerlach V,  Shimkets RA,  Malyankar UM;
PI   Boldog FL,  Zerhusen BD,  Spytek KA,  Majumder K,  Tchernev VT;
PI   Padigaru M,  Patturajan M,  Burgess CE,  Gangolli EA,  Smithson G;
PI   Rastelli L,  Macdougall JR,  Taupier RJ,  Grosse WM,  Szekeres ES;
PI   Alsobrook JP;
XX
DR   WPI; 2002-049278/06.
DR   P-PSDB; AAM47672.
XX
PT   Novel G-protein coupled receptor-related polypeptides and
PT   polynucleotides for diagnosing, preventing and treating cardiomyopathy,
PT   atherosclerosis, disorders related to cell signal processing and for
PT   identifying modulators  -
XX
PS   Claim 8; Page 76; 227pp; English.
XX
CC   The present invention relates to novel G-coupled protein-receptor related
CC   proteins and coding sequences (MOLX, where X is a number from 1 to 10,
CC   ABA04589-ABA04603 and AAM47659-AAM47673). MOLX proteins and coding
CC   sequences are useful for treating or preventing a MOLX-associated
CC   disorder, such as cardiomyopathy, atherosclerosis, disorders related to
CC   cell signal processing and metabolic pathway modulation, diabetes and
CC   cancer. Additionally, MOLX proteins and coding sequences are useful for
CC   preventing and treating a variety of disorders including metabolic
CC   disorders, nutritional oedema, chronic and hereditary pancreatitis,
CC   obesity, infectious disease, anorexia, neurodegenerative disorders,
CC   Alzheimer's disease, Parkinson's disease, stroke, immune disorders,
CC   haematopoietic disorders and various dyslipidaemias, metabolic syndrome X
CC   and wasting disorders associated with chronic diseases and cancers,
CC   cardiac disorders, hypertension, hypercalcaemia, cirrhosis, angiogenesis
CC   and wound healing, trauma, glomerulonephritis, hyper and hypothyroidism,
CC   multiple sclerosis, lung diseases including asthma, Crohn's disease,
CC   scleroderma, autoimmune diseases, developmental disorders and neural tube
CC   defects. The present sequence is the coding sequence for MOL10a.
CC   MOL10a protein is a cyclic nucleotide-gated olfactory channel-like
CC   protein.
XX
SQ   Sequence 1835 BP; 420 A; 521 C; 521 G; 373 T; 0 other;
```

```
  Query Match             97.7%;  Score 1687.4;  DB 24;  Length 1835;
  Best Local Similarity   99.4%;  Pred. No. 0;
  Matches 1727;  Conservative    0;  Mismatches    1;  Indels    9;  Gaps    3;

Qy       1 atgagccaggacaccaaagtgaagacaacagagtccagtcccccagccccatccaaggcc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      54 atgagccaggacaccaaagtgaagacaacagagtccagtcccccagccccatccaaggcc 113

Qy      61 ---aggaagttgctgcctgtcctggacccatctggggattactactactggtggctgaac 117
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     114 aggaggaagttgctgcctgtcctggacccatctggggattactactactggtggctgaac 173

Qy     118 acaatggtcttcccagtcatgtataacctcatcatcctcgtgtgcagagcctgcttcccc 177
           |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db     174 acaatggtcttcccagtcatgtataacctcatcatcctcgtgtgcagggcctgcttcccc 233

Qy     178 gacttgcagcacggttatctggtggcctggttggtgctggactacacgagtgacctgcta 237
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     234 gacttgcagcacggttatctggtggcctggttggtgctggactacacgagtgacctgcta 293

Qy     238 tacctactagacatggtggtgcgcttccacaca---ggattcttggaacagggcatcctg 294
           ||||||||||||||||||||||||||||||||   |||||||||||||||||||||||||
Db     294 tacctactagacatggtggtgcgcttccacacaggtggattcttggaacagggcatcctg 353

Qy     295 gtggtggacaagggtaggatctcgagtcgctacgttcgcacctggagtttcttcttggac 354
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     354 gtggtggacaagggtaggatctcgagtcgctacgttcgcacctggagtttcttcttggac 413

Qy     355 ctggcttccctgatgcccacagatgtggtctacgtgcggctgggcccgcacacacccacc 414
```

```
Db    414 ctggcttccctgatgcccacagatgtggtctacgtgcggctgggcccgcacacacccacc 473

Qy    415 ctgaggctgaaccgctttctccgcgcgccccgcctcttcgaggccttcgaccgcacagag 474
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    474 ctgaggctgaaccgctttctccgcgcgccccgcctcttcgaggccttcgaccgcacagag 533

Qy    475 acccgcacagcttacccaaatgcctttcgcattgccaagctgatgctttacatttttgtc 534
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    534 acccgcacagcttacccaaatgcctttcgcattgccaagctgatgctttacatttttgtc 593

Qy    535 gtcatccattggaacagctgcctatactttgccctatcccggtacctgggcttcgggcgt 594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    594 gtcatccattggaacagctgcctatactttgccctatcccggtacctgggcttcgggcgt 653

Qy    595 gacgcatgggtgtacccggaccccgcgcagcctggctttgagcgcctgcggcgccagtac 654
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    654 gacgcatgggtgtacccggaccccgcgcagcctggctttgagcgcctgcggcgccagtac 713

Qy    655 ctctatagcttttacttctccacgctgatactgactacagtgggcgatacaccgccgcca 714
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    714 ctctatagcttttacttctccacgctgatactgactacagtgggcgatacaccgccgcca 773

Qy    715 gccagggaagaagagtacctcttcatggtgggcgacttcctgctggccgtcatgggtttc 774
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    774 gccagggaagaagagtacctcttcatggtgggcgacttcctgctggccgtcatgggtttc 833

Qy    775 gccaccatcatgggtagcatgagctctgtcatctacaacatgaacactgcagatgcggct 834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    834 gccaccatcatgggtagcatgagctctgtcatctacaacatgaacactgcagatgcggct 893

Qy    835 ttctacccagatcatgcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaag 894
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    894 ttctacccagatcatgcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaag 953

Qy    895 ctggagcggcgagttattgactggtatcagcacctgcagatcaacaagaagatgaccaac 954
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    954 ctggagcggcgagttattgactggtatcagcacctgcagatcaacaagaagatgaccaac 1013

Qy    955 gaggtagccatcttacagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtg 1014
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1014 gaggtagccatcttacagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtg 1073

Qy   1015 cacctgtccactctgagccgggtgcagatctttcagaactgtgaggccagcctgctggag 1074
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1074 cacctgtccactctgagccgggtgcagatctttcagaactgtgaggccagcctgctggag 1133

Qy   1075 gagctggtgctgaagctgcagccccagacctactcaccaggtgaatatgtatgccgcaaa 1134
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1134 gagctggtgctgaagctgcagccccagacctactcaccaggtgaatatgtatgccgcaaa 1193

Qy   1135 ggagacattggccaagagatgtacatcatccgagagggtcaactggccgtggtggcagat 1194
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1194 ggagacattggccaagagatgtacatcatccgagagggtcaactggccgtggtggcagat 1253

Qy   1195 gatggtatcacacagtatgctgtgctcggtgcagggctctactttggggagatcagcatc 1254
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1254 gatggtatcacacagtatgctgtgctcggtgcagggctctactttggggagatcagcatc 1313

Qy   1255 atcaacatcaaa---gggaacatgtctgggaaccgccgcacagccaacatcaagagccta 1311
          ||||||||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db   1314 atcaacatcaaaggtgggaacatgtctgggaaccgccgcacagccaacatcaagagccta 1373

Qy   1312 ggttattcagacctattctgcctgagcaaggaggacctgcgggaggtgctgagcgagtat 1371
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1374 ggttattcagacctattctgcctgagcaaggaggacctgcgggaggtgctgagcgagtat 1433

Qy   1372 ccacaagcacagaccatcatggaggagaaaggacgtgagatcctgctgaaaatgaacaag 1431
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1434 ccacaagcacagaccatcatggaggagaaaggacgtgagatcctgctgaaaatgaacaag 1493

Qy   1432 ttggacgtgaatgctgaggcagctgagatcgccctgcaggaggccacagagtcccggcta 1491
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1494 ttggacgtgaatgctgaggcagctgagatcgccctgcaggaggccacagagtcccggcta 1553

Qy   1492 cgaggcctagaccagcagctggatgatctacagaccaagtttgctcgcctcctggctgag 1551
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1554 cgaggcctagaccagcagctggatgatctacagaccaagtttgctcgcctcctggctgag 1613

Qy   1552 ctggagtccagcgcacttaagattgcttaccgcattgaacggctggagtggcagactcga 1611
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1614 ctggagtccagcgcacttaagattgcttaccgcattgaacggctggagtggcagactcga 1673

Qy   1612 gagtggccaatgcccgaggacctggctgaggctgatgacgagggtgagcctgaggaggga 1671
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1674 gagtggccaatgcccgaggacctggctgaggctgatgacgagggtgagcctgaggaggga 1733

Qy   1672 acttccaaagatgaagagggcagggccagccaggagggaccccccaggtccagagtga 1728
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1734 acttccaaagatgaagagggcagggccagccaggagggaccccccaggtccagagtga 1790
```

```
RESULT   3
AAD09558
ID   AAD09558 standard; cDNA; 1995 BP.
XX
AC   AAD09558;
XX
DT   10-SEP-2001  (first entry)
XX
DE   Human transporter and ion channel-7 (TRICH-7) cDNA.
XX
KW   Human; transporter and ion channel-7; TRICH-7; vaccine; cystic fibrosis;
KW   gene therapy; amyotrophic lateral sclerosis; amnesia; muscular dystrophy;
KW   hypertension; angina; neurological disorder; asthma; bipolar disorder;
KW   dementia; depression; Alzheimer's disease; epilepsy; mood; arrhythmia;
KW   Pick's disease; ischaemic cerebrovascular disease; AIDS; anxiety; stroke;
KW   Huntington's disease; Parkinson's disease; cerebral neoplasm; allergy;
KW   demyelinating disease; mental disorder; Schizophrenia; polymyositis;
KW   muscle disorder; cardiomyopathy; cataract; myocarditis; Grave's disease;
KW   dermatomyositis; diabetes mellitus; immunological disorder; psoriasis;
KW   rheumatoid arthritis; Sjogren's syndrome; systemic lupus erythematosus;
KW   sickle cell anaemia; Wilson's disease; infertility; Cushing's disease;
KW   scleroderma; pulmonary artery stenosis; nootropic; Addison's disease;
KW   malabsorption syndrome; hypercholesterolaemia; cancer; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1995
FT                   /*tag= a
FT                   /product= "Human TRICH-7 protein"
XX
PN   WO200146258-A2.
XX
PD   28-JUN-2001.
XX
PF   22-DEC-2000; 2000WO-US35095.
XX
PR   23-DEC-1999;   99US-0172000.
PR   14-JAN-2000; 2000US-0176083.
PR   21-JAN-2000; 2000US-0177332.
PR   28-JAN-2000; 2000US-0178572.
PR   02-FEB-2000; 2000US-0179758.
PR   10-FEB-2000; 2000US-0181625.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Baughn MR,  Burford N,  Au-Young J,  Lu DAM,  Yang J,  Reddy R,  Lal P;
PI   Hillman JL,  Azimzai Y,  Yue H,  Nguyen DB,  Yao MG,  Gandhi AR;
PI   Tang YT,  Khan FA;
XX
DR   WPI; 2001-418042/44.
DR   P-PSDB; AAE04894.
XX
PT   Novel human transporter and ion channel proteins useful for treating
```

PT and preventing transport, neurological, muscle and immunological
PT disorders -
XX
PS Claim 5; Page 145; 160pp; English.
XX
CC The present sequence is transporter and ion channel-7 (TRICH-7) cDNA.
CC TRICH is used as vaccine. TRICH is useful for treating a disease or
CC condition associated with decreased expression of functional TRICH,
CC such as transport disorder including amyotrophic lateral sclerosis,
CC cystic fibrosis, Becker's muscular dystrophy, Charcot-Marie Tooth
CC disease, Duchenne muscular dystrophy, angina and hypertension,
CC neurological disorders including Alzheimer's disease, amnesia, bipolar
CC disorder, dementia, depression, epilepsy, ischaemic cerebrovascular
CC disease, stroke, cerebral neoplasms, Pick's disease, Huntington's
CC disease and Parkinson's disease, demyelinating diseases, mental disorders
CC including mood, anxiety, Schizophrenia and seasonal affective disorder,
CC muscle disorder including cardiomyopathy, myocarditis, polymyositis,
CC dermatomyositis, arrhythmias and asthma and immunological disorders
CC including AIDS, adult respiratory distress syndrome (ARDS), allergies,
CC anaemia, diabetes mellitus, rheumatoid arthritis, scleroderma, Sjogren's
CC syndrome, systemic lupus erythematosus and other diseases including
CC sickle cell anaemia, Wilson's disease, cataracts, infertility, pulmonary
CC artery stenosis, Grave's disease, Cushing's disease, Addison's disease,
CC glucose-galactose malabsorption syndrome, hypercholesterolaemia, cancers
CC psoriasis and viral, bacterial, fungal, helminthic and protozoal
CC infections. TRICH DNA is useful in gene therapy and in diagnostic
CC purposes.
XX
SQ Sequence 1995 BP; 511 A; 511 C; 532 G; 441 T; 0 other;

Query Match 33.5%; Score 578.6; DB 22; Length 1995;
Best Local Similarity 60.7%; Pred. No. 1.4e-144;
Matches 966; Conservative 0; Mismatches 619; Indels 6; Gaps 1;

```
Qy     70 ctgcctgtcctggacccatctggggattactactactggtggctgaacacaatggtcttc 129
Db    388 ctatttgtcttggacccagctggggattggtactactgctggctatttgtcattgccatg 447

Qy    130 ccagtcatgtataacctcatcatcctcgtgtgcagagcctgcttccccgacttgcagcac 189
Db    448 cccgtcctttacaactggtgcctgctggtggccagagcctgcttcagtgacctacagaaa 507

Qy    190 ggttatctggtggcctggttggtgctggactacacgagtgacctgctatacctactagac 249
Db    508 ggctactacctggtgtggctggtgctggattatgtctcagatgtggtctacattgcggac 567

Qy    250 atggtggtgcgcttccacacaggattcttggaacagggcatcctggtggtggacaagggt 309
Db    568 ctcttcatccgattgcgcacaggtttcctggagcaggggctgctggtcaaagataccaag 627

Qy    310 aggatctcgagtcgctacgttcgcacctggagtttcttcttggacctggcttccctgatg 369
Db    628 aaactgcgagacaactacatccacaccctgcagttcaagctggatgtggcttccatcatc 687

Qy    370 cccacagatgtggtctacgtgcggctgggcccgcacacacccaccctgaggctgaaccgc 429
Db    688 cccactgacctgatctattttgctgtggacatccacagccctgaggtgcgcttcaaccgc 747

Qy    430 tttctccgcgcgccccgcctcttcgaggccttcgaccgcacagagacccgcacagcttac 489
Db    748 ctgctgcactttgcccgcatgtttgagttctttgaccggacagagacacgcaccaactac 807

Qy    490 ccaaatgcctttcgcattgccaagctgatgctttacatttttgtcgtcatccattggaac 549
Db    808 cctaacatcttccgcatcagcaaccttgtcctctacatcttggtcatcatccactggaat 867

Qy    550 agctgcctatactttgccctatcccggtacctgggcttcgggcgtgacgcatgggtgtac 609
Db    868 gcctgcatctattatgccatctccaaatccataggctttggggtcgacacctgggtttac 927

Qy    610 ccggaccccgcgcagcctggctttgagcgcctgcggcgccagtacctctatagcttttac 669
Db    928 ccaaacatcactgaccctgagtatggctacctggctagggaatacatctattgcctttac 987

Qy    670 ttctccacgctgatactgactacagtgggcgatacaccgccgccagccagggaagaagag 729
Db    988 tggtccacactgactctcactaccattggggagacaccacccccctgtaaaggatgaggag 1047

Qy    730 tacctcttcatggtgggcgacttcctgctggccgtcatgggtttcgccaccatcatgggt 789
Db   1048 tacctatttgtcatctttgacttcctgattggcgtcctcatctttgccaccatcgtggga 1107

Qy    790 agcatgagctctgtcatctacaacatgaacactgcagatgcggctttctacccagatcat 849
Db   1108 aatgtgggctccatgatctccaacatgaatgccacccgggcagagttccaggctaagatc 1167

Qy    850 gcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggagcggcgagtt 909
Db   1168 gatgccgtgaaacactacatgcagttccgaaaggtcagcaaggggatggaagccaaggtc 1227

Qy    910 attgactggtatcagcacctgcagatcaacaagaagatgaccaacgaggtagccatctta 969
Db   1228 attaggtggtttgactacttgtggaccaataagaagacagtggatgagcgagaaattctc 1287

Qy    970 cagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctgtccactctg 1029
Db   1288 aagaatctgccagccaagctcagggctgagatagccatcaatgtccacttgtccacactc 1347

Qy   1030 agccgggtgcagatctttcagaactgtgaggccagcctgctggaggagctggtgctgaag 1089
Db   1348 aagaaagtgcgcatcttccatgattgtgaggctggcctgctggtagagctggtactgaaa 1407

Qy   1090 ctgcagccccagacctactcaccaggtgaatatgtatgccgcaaaggagacattggccaa 1149
Db   1408 ctccgtcctcaggtcttcagtcctggggattacatttgccgcaaaggggacatcggcaag 1467

Qy   1150 gagatgtacatcatccgagagggtcaactggccgtggtggcagatgatggtatcacacag 1209
Db   1468 gagatgtacatcattaaggagggcaaactggcagtggtggctgatgatggtgtgactcag 1527

Qy   1210 tatgctgtgctcggtgcagggctctactttggggagatcagcatcatcaacatcaaaggg 1269
Db   1528 tatgctctgctgtcggctggaagctgctttggcgagatcagtatccttaacattaagggc 1587

Qy   1270 aacatgtctgggaaccgccgcacagccaacatcaagagcctaggttattcagacctattc 1329
Db   1588 agtaaaatgggcaatcgacgcacagctaatatccgcagcctgggctactcagatctcttc 1647

Qy   1330 tgcctgagcaaggaggacctgcgggaggtgctgagcgagtatccacaagcacagaccatc 1389
Db   1648 tgcttgtccaaggatgatcttatggaagctgtgactgagtaccctgatgccaagaaagtc 1707

Qy   1390 atggaggagaaaggacgtgagatcctgctgaaaatgaacaagttggacgtgaatgctgag 1449
Db   1708 ctagaagagaggggtcgggagatcctcatgaaggagggactgctggatgagaacgaagtg 1767

Qy   1450 gcagctgagatcgccctgcaggaggccacagagtcccggctacgaggcctagaccagcag 1509
Db   1768 gca------accagcatggaggtcgacgtgcaggagaagctagggcagctggagaccaac 1821

Qy   1510 ctggatgatctacagaccaagtttgctcgcctcctggctgagctggagtccagcgcactt 1569
Db   1822 atggaaaccttgtacactcgctttggccgcctgctggctgagtacacgggggcccagcag 1881

Qy   1570 aagattgcttaccgcattgaacggctggagtggcagactcgagagtggccaatgcccgag 1629
Db   1882 aagctcaagcagcgcatcacagttctggaaaccaagatgaaacagaacaatgaagatgac 1941

Qy   1630 gacctggctgaggctgatgacgagggtgagc 1660
Db   1942 tacctgtctgatgggatgaacagccctgagc 1972
```

RESULT 4
AAH98838

```
ID   AAH98838 standard; cDNA; 995 BP.
XX
AC   AAH98838;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Rat EST-derived coding sequence SEQ ID NO: 695.
XX
KW   Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW   tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW   diagnostics; forensic test; gene mapping; genetic disorder;
KW   biodiversity; gene therapy; nutrition; ss.
XX
OS   Rattus norvegicus.
XX
PN   WO200154477-A2.
XX
PD   02-AUG-2001.
XX
PF   25-JAN-2001; 2001WO-US02687.
XX
PR   25-JAN-2000; 2000US-0491404.
PR   17-JUL-2000; 2000US-0617746.
PR   03-AUG-2000; 2000US-0631451.
PR   15-SEP-2000; 2000US-0663870.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Zhou P,  Qian XB,  Wang Z,  Chen R,  Asundi V;
PI   Cao Y,  Drmanac RA,  Zhang J,  Werhman T;
XX
DR   WPI; 2001-476164/51.
DR   P-PSDB; AAM24179.
XX
PT   Isolated polypeptide for treatment of diseases, diagnostics, raising
PT   antibodies and research use -
XX
PS   Claim 1; Page 634; 1275pp; English.
XX
CC   The present invention provides the protein and coding sequences of novel
CC   proteins from a variety of organisms, including human, dog, cat, horse,
CC   cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC   urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC   from the organism of interest. They can be used in diagnostics,
CC   forensics, gene mapping, identification of mutations, to assess
CC   biodiversity and for nutritional purposes. The present sequence is a cDNA
CC   of the invention.
XX
SQ   Sequence 995 BP; 187 A; 291 C; 269 G; 248 T; 0 other;


  Query Match             30.6%;  Score 528.6;  DB 22;  Length 995;
  Best Local Similarity   97.3%;  Pred. No. 2.5e-131;
  Matches  548;  Conservative    0;  Mismatches   14;  Indels    1;  Gaps    1;

Qy      61 aggaagttgctgcctgtcctggacccatctggggattactactactggtggctgaacaca 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 aggaagttgctgcctgtcctggacccatctggggattactactactggtggctgaacaca 180

Qy     121 atggtcttcccagtcatgtataacctcatcatcctcgtgtgcagagcctgcttccccgac 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 atggtcttcccagtcatgtataacctcatcatcctcgtgtgcagagcctgcttccccgac 240

Qy     181 ttgcagcacggttatctggtggcctggttggtgctggactacacgagtgacctgctatac 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 ttgcagcacggttatctggtggcctggttggtgctggactacacgagtgacctgctatac 300

Qy     241 ctactagacatggtggtgcgcttccacacaggattcttggaacagggcatcctggtggtg 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ctactagacatggtggtgcgcttccacacaggattcttggaacagggcatcctggtggtg 360

Qy     301 gacaagggtaggatctcgagtcgctacgttcgcacctggagtttcttcttggacctggct 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 gacaagggtaggatctcgagtcgctacgttcgcacctggagtttcttcttggacctggct 420

Qy     361 tccctgatgcccacagatgtggtctacgtgcggctgggcccgcacacacccaccctgagg 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 tccctgatgcccacagatgtggtctacgtgcggctgggcccgcacacacccaccctgagg 480

Qy     421 ctgaaccgctttctccgcgcgccccgcctcttcgaggccttcgaccgcacagagacccgc 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ctgaaccgctttctccgcgcgccccgcctcttcgaggccttcgaccgcacagagacccgc 540

Qy     481 acagcttacccaaatgcctttcgcattgccaagctgatgctttacatttttgtcgtcatc 540
           |||||||||||||||||||| |||||| |||||||||||||||||||||| |  ||||
Db     541 acagcttacccaaatgcctttgcattggcaagctgatgctttacatttttggccgcatc 600

Qy     541 cattggaacagctgcctatactttgccctatcccggtacctgggcttcgggcgtg-acgc 599
           |||||||||| |||||||||||| |||||||||||||||||||| ||||||| || |
Db     601 cattggaacaactgcctatacttttccctatcccggtacctgggctttgggcgtgaaccc 660

Qy     600 atgggtgtacccggaccccgcgc 622
           |||||||||||| |  |||  ||
Db     661 atgggtgtaccccggaccccggc 683


RESULT   5
ABL13353
ID   ABL13353 standard; cDNA; 2909 BP.
XX
AC   ABL13353;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 34541.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB69250.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 34541; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2909 BP; 729 A; 726 C; 804 G; 650 T; 0 other;
```

Query Match 23.4%; Score 405.2; DB 23; Length 2909;
Best Local Similarity 55.1%; Pred. No. 4.3e-98;
Matches 856; Conservative 0; Mismatches 668; Indels 30; Gaps 2;

```
Qy      78 cctggacccatctggggattactactactggtggctgaacacaatggtcttcccagtcat 137
Db     421 cgtcgatcccaccctgcagtcgcattacagatggctggccatcgtttcactggcagtgct 480

Qy     138 gtataacctcatcatcctcgtgtgcagagcctgcttccccgacttgcagcacggttatct 197
Db     481 ttacaatatcatcttcgtggtgggacgtgcggtcttctgggagatcaataagagtgctcc 540

Qy     198 ggtggcctggttggtgctggactacacgagtgacctgctatacctactagacatggtggt 257
Db     541 cgccttttggtacacgctggactacctgtgcgactttatctatctgctggatacgctcgt 600

Qy     258 gcgcttccacacaggattcttggaacagggcatcctggtggtggacaagggtaggatctc 317
Db     601 ccacatgcacgagggatttttggaccaaggtctcctcgtgcgggatgcctttcggctgcg 660

Qy     318 gagtcgctacgttcgcacctggagtttcttcttggacctggcttccctgatgcccacaga 377
Db     661 ccggcactatttccacaccaagggctggtacctcgacgtcctgtccatgctgcccacgga 720

Qy     378 tgtggtctacgtgcggctgggcccgcacacacccacc---------------------- 414
Db     721 cttggcctacatctggtggccgccggagacctgctccagtctgtatctgccgtgtcccgt 780

Qy     415 ----ctgaggctgaaccgctttctccgcgcgccccgcctcttcgaggccttcgaccgcac 470
Db     781 gatagtgcgtctgaatcgtctgctcaggatcaatcgactgtgggagtggttcgatcgaac 840

Qy     471 agagacccgcacagcttacccaaatgcctttcgcattgccaagctgatgctttacatttt 530
Db     841 ggagacagcgacgggatatccgaatgccttccgcatctgcaaggtggtgctggccatttt 900

Qy     531 tgtcgtcatccattggaacagctgcctatactttgccctatcccggtacctgggcttcgg 590
Db     901 ggtgctgatccactggaatgcgtgcatgtactttgccataagctacgaaatcggcttcag 960

Qy     591 gcgtgacgcatgggtgtacccggaccccgcgcagcctggctttgagcgcctgcggcgcca 650
Db     961 ttcggactcgtgggtgta---taatctgaatggaacgaggaacaatacgttgcagcggca 1017

Qy     651 gtacctctatagcttttacttctccacgctgatactgactacagtgggcgatacaccgcc 710
Db    1018 atatatctatagtttctactggtcgacgctaacgctgaccaccattggcgagactccaac 1077

Qy     711 gccagccagggaagaagagtacctcttcatggtgggcgacttcctgctggccgtcatggg 770
Db    1078 gccggagaacgatgtggagtatctgttcgtggtggccgatttcctggctggcgtcctaat 1137

Qy     771 tttcgccaccatcatgggtagcatgagctctgtcatctacaacatgaacactgcagatgc 830
Db    1138 cttcgccaccattgtgggtaacattggctccatgatctccaacatgaatgtggcccgcgt 1197

Qy     831 ggctttctacccagatcatgcactggtgaagaagtacatgaagctgcagcacgtcaaccg 890
Db    1198 ggagttccagaatcgcatggacggagtcaagcagtacatggccttccggcgagtgggtca 1257

Qy     891 caagctggagcggcgagttattgactggtatcagcacctgcagatcaacaagaagatgac 950
Db    1258 cgagctggaggcccgggtgatccgctggtttgcctacacgtggtcgcagagtggtgcgct 1317

Qy     951 caacgaggtagccatcttacagcacttgcctgagcggctgcgggcagaagtggctgtgtc 1010
Db    1318 ggacgaggagcgtgtgctggccgctctgccggacaagctgaaggcggagatcgctatcca 1377

Qy    1011 tgtgcacctgtccactctgagccgggtgcagatctttcagaactgtgaggccagcctgct 1070
Db    1378 ggtgcacatggatacgctgaagcaggtgcgcatcttccacgacacggaaccgggtctcct 1437

Qy    1071 ggaggagctggtgctgaagctgcagccccagacctactcaccaggtgaatatgtatgccg 1130
Db    1438 cgaggccctggtgctgaagctcaaactgcaggtcttcagtccgggcgactatatctgtcg 1497

Qy    1131 caaaggagacattggccaagagatgtacatcatccgagagggtcaactggccgtggtggc 1190
Db    1498 caagggtgacgtgggcaaggagatgtacatcgtgaagcgtggcaagttgtccgtcgtcgg 1557

Qy    1191 agatgatggtatcacacagtatgctgtgctcggtgcagggctctactttggggagatcag 1250
Db    1558 tgatgatggcatcaccgtgctggccaccttgggcgccggttcggtgttcggcgaggtttc 1617

Qy    1251 catcatcaacatcaaagggaacatgtctgggaaccgccgcacagccaacatcaagagcct 1310
Db    1618 cgtactcgagattgcgggcaatcggacgggcaatcggcgcaccgccaatgtgagatccct 1677

Qy    1311 aggttattcagacctattctgcctgagcaaggaggacctgcgggaggtgctgagcgagta 1370
Db    1678 gggctactcggatctcttctgtctggcgaaacgggatttgtgggagaccctgtccgacta 1737

Qy    1371 tccacaagcacagaccatcatggaggagaaaggacgtgagatcctgctgaaaatgaacaa 1430
Db    1738 tccggaggccagatccacgctcactcagcgaggatgccagctgctgcgcaaggatggctt 1797

Qy    1431 gttggacgtgaatgctgaggcagctgagatcgccctgcaggaggccacagagtcccggct 1490
Db    1798 actggacgagcaaatattcgctgattcacaaagggtacacgacagcatcgagggcggcat 1857

Qy    1491 acgaggcctagaccagcagctggatgatctacagaccaagtttgctcgcctcctggctga 1550
Db    1858 cgaaaagctggagctctcggtggagaacctcaacatgcggctggcccgcctcctggctga 1917

Qy    1551 gctggagtccagcgcacttaagattgcttaccgcattgaacggctggagtggca 1604
Db    1918 gtacacggcaagccaggccaagatcaagcagcgcctcgcgaagctggagatgaa 1971
```

RESULT 6
ABL18085
ID ABL18085 standard; DNA; 2109 BP.
XX
AC ABL18085;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 5728.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 5728; 21pp + Sequence Listing; English.

```
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2109 BP; 485 A; 589 C; 605 G; 430 T; 0 other;


  Query Match             16.2%;  Score 279.2;  DB 23;  Length 2109;
  Best Local Similarity   52.6%;  Pred. No. 2e-64;
  Matches  689;  Conservative    0;  Mismatches  603;  Indels   18;  Gaps    3;

Qy    122 tggtcttcccagtcatgtataacctcatcatcctcgtgtgcagagcctgcttccccgact 181
          ||  | |  |  || |||| ||| ||    |  |  | | | |  |  |||||  ||
Db      8 tgtccatggccttcctgtacaacttctgggtgataatctaccgcttcgccttccaggaga 67

Qy    182 tgcagcacggttatctggtggcctggttggtgctggactacacgagtgacctgctatacc 241
          |  | |   |     | |     |||||    |||||||||  |   ||| | || || |
Db     68 tcaatcggcgaacgatcgccatttggttctgtctggactacctgtccgacttcctgtatc 127

Qy    242 tactagacatggtggtgcgcttccacacaggattcttggaacagggcatcctggtggtgg 301
          |  | || ||  || | | ||| | ||| |||| | ||||  | ||  | |||  |  ||
Db    128 tgatcgatatactgttccactttcgcaccggatacctggaggacggggtgctgcagacgg 187

Qy    302 acaagggtaggatctcgagtcgctacgttcgcacctggagtttcttcttggacctggctt 361
          ||      | | |    |  | |||| |  | |   ||  |||| | | |||     |
Db    188 acgctttgaagctgcgcacccactacatgaactcgacgatcttctacatcgactgcctgt 247

Qy    362 ccctgatgcccacagatgtggtctacgtgcggctgggcccgcacacacccaccctgaggc 421
           |||| ||||    ||  |  |||| ||    | |||    ||    | | ||||  |
Db    248 gcctgctgccgctggacttttctctatttgtccatcggctttaac---tcgatcctgcgca 304

Qy    422 tgaaccgctttctccgcgcgccccgcctcttcgaggccttcgaccgcacagagacccgca 481
              |||  |  |          |||  |||  |    | | || |||||  |||   | ||
Db    305 gcttccggctggtgaagatctaccggttctgggccttcatggatcgcaccgagcggcaca 364

Qy    482 cagcttacccaaatgcctttcgcattgccaagctgatgctttacatttttgtcgtcatcc 541
          |    ||||| ||    || || |    |    || || |  ||| |  ||||  |  |||
Db    365 ccaactacccgaacctgttccggagcacggccctcatccactacctgcttgtgatattcc 424

Qy    542 attggaacagctgcctatactttgccctatcccggtacctgggcttcgggcgtgacgcat 601
          ||||||||  | || || |||     | ||  |  | ||   || |||||      |   |
Db    425 attggaacggttgtctctaccacatcatacacaagaacaatgggttcggatcacgcaact 484

Qy    602 gggtgtacccggaccccgcgcagcctggctttgagcgcctgcggcgccagtacctctata 661
          ||||  ||||  ||| | | |  | | | | | |               |||||  ||  | |
Db    485 gggtctaccacgactcggagtcggcggacgtggtcaag---------cagtatctgcaga 535

Qy    662 gcttttacttctccacgctgatactgactacagtgggcgatacaccgccgccagccaggg 721
          |||  ||||  | ||| |||   ||||| ||  | || |||   ||   |||  |
Db    536 gctactactggtgcaccctggccctgaccaccatcggggatcttcccaagccgcgctcca 595

Qy    722 aagaagagtacctcttcatggtgggcgacttcctgctggccgtcatgggtttcgccacca 781
          | |  |||||| |||| |  |     |  | ||  | | | | |||   ||||||||
Db    596 agggcgagtacgtctttgttatcctgcagctgctctttggcctgatgctcttcgccacgg 655

Qy    782 tcatgggtagcatgagctctgtcatctacaacatgaacactgcagatgcggctttctacc 841
          |  ||||   | ||  |    |  |        |||   | ||      ||  ||  |
Db    656 tactggggcacgtggccaacattgtgacgtcagtgagtgcagcgcgcaaggagtttcaag 715

Qy    842 cagatcatgcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggagc 901
          |  | |  |     ||||||| |||||||  | |||  |  ||        | ||| ||
Db    716 ccaagctggatggcgtgaagacgtacatgcggatgcgacgtgtgccgaatcatctgcagg 775

Qy    902 ggcgagttattgactggtatcagcacctgcagatcaacaagaagatgaccaacgaggtag 961
```

```
              |   || ||  | ||||   |  |||||  | | |   | ||     |  ||||||
Db    776 tgaaggtcatcaaatggttcgattacctgtggcttacgcaaaaatgctcggacgaggagc 835

Qy    962 ccatcttacagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctgt 1021
          |  | |        | |||||     |   ||| ||| | ||  |    || ||  |
Db    836 gcgccgtgtcctgtcttcctgataaattaaaggctgaaatagcaattaacgtccatttag 895

Qy    1022 ccactctgagccgggtgcagatctttcagaactgtgaggccagcctgctggaggagctgg 1081
           || || |  |||||| |||| || || |||   |||||| || | ||    || ||||
Db     896 atacgctcaagcgggtggagattttccaaaacacagaggccggcttcctatgcgaactgg 955

Qy    1082 tgctgaagctgcagccccagacctactcaccaggtgaatatgtatgccgcaaaggagaca 1141
           |||||   |||  ||||  |  || ||| || || || ||  | ||| | || || ||
Db     956 tgctgcgcctgaggcccgtgctcttctcgcccggcgactacatctgcagaaagggcgagg 1015

Qy    1142 ttggccaagagatgtacatcatccgagagggtcaactggccgtggtggcagatgatggta 1201
           | ||| | ||||||||||||  |       || | | ||   |||||||| ||  |||| |
Db    1016 tgggcaaggagatgtacattgtgaaccgaggacgattgcaggtggtggccgacaatggaa 1075

Qy    1202 tcacacagtatgctgtgctcggtgcagggctctactttggggagatcagcatcatcaaca 1261
           ||   |  ||   ||    || ||  ||| ||||| ||||| || ||  |||| |
Db    1076 agacggtgatggcctccctgaaggctggttcctattttggcgagattagtatactcaata 1135

Qy    1262 tcaaagggaacatgtctgggaaccgccgcacagccaacatcaagagcctaggttattcag 1321
           |      |  ||   | || ||||| |||||||||| | |         | || ||    |
Db    1136 t------gggcaccgcaggcaaccgacgcacagccagcgttcgctcagtgggatacagcg 1189

Qy    1322 acctattctgcctgagcaaggaggacctgcgggaggtgctgagcgagtatccacaagcac 1381
           |||| |||  |||||||||| ||||| || |||| |||||||  ||||||||    || |
Db    1190 acctcttcgtcctgagcaagaaggacatgtgggacgtgctgaaggagtatccggcggcgc 1249

Qy    1382 agaccatcatggaggagaaaggacgtgagatcctgctgaaaatgaacaag 1431
                   |||||  || ||      ||    ||  ||||   || |||
Db    1250 gtgttcgtctggagtcgatagccgtcaagcgattggagaaatacaagaag 1299


RESULT   7
ABL14931
ID   ABL14931 standard; cDNA; 5133 BP.
XX
AC   ABL14931;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 39275.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB70828.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 39275; 21pp + Sequence Listing; English.
XX
```

```
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 5133 BP; 1201 A; 1449 C; 1501 G; 982 T; 0 other;


  Query Match           12.9%;  Score 222.2;  DB 23;  Length 5133;
  Best Local Similarity   50.7%;  Pred. No. 5.3e-49;
  Matches  713;  Conservative    0;  Mismatches  653;  Indels   39;  Gaps    6;

Qy      83 acccatctggggattactactactggtggctgaacacaatggtcttcccagtcatgtata 142
           | |||  ||   |||  |||| ||  ||| | |   |    |     ||  | ||||
Db       5 atccagatgaaaatttttacttctattggttaatgatgttaactgtatgtgttctatata 64

Qy     143 acctcatcatcctcgtgtgcagagcctgcttccccgacttgcagcacggttatctggtgg 202
           | ||    | |||  |    ||     |||| || ||  |||||||   |   |    |
Db      65 atctatggaccctaattgtgaggcagagctttcctgaactgcagcaatctgtgcccacgt 124

Qy     203 cctggttggtgctggactacacgagtgacctgctatacctactagacatggtggtgcgct 262
             ||| |  |    || |  | ||  ||  |  |||   | ||||| ||  | ||  |  |
Db     125 tttggctcatctgcgattcgatgacagatgttgtatttatcctagatataatagttcaat 184

Qy     263 tccacacaggattcttggaacagggcatcctggtggtggacaagggtaggatctcgagtc 322
           | | ||||||  |   |||| ||||||  |  ||||    |||  |  | | |  |  |||
Db     185 tacgcacaggctatctggagcagggcctaatggtatacgacgacaggaagctggcctgtc 244

Qy     323 gctacgttcgcacctggagtttcttcttggacctggcttccctgatgcccacagatgtgg 382
            ||||||||  | |  |    ||  |||| ||  ||    | |||||  ||    |||  ||
Db     245 actacgttcactcgcgcgactttatcttcgatatgatagcgctgataccattggatttgc 304

Qy     383 tctacgtgcggctgggcccgcacacacccaccctgaggctgaaccgctttctccgcgcgc 442
           |  |  |   | |||||  | ||  |             |||    || |||   ||||
Db     305 tgcagctcaagatgggcacacatcc------------gctcttgcgttttacgcgcgttt 352

Qy     443 cccgcctcttcgaggccttcgaccgcacagagacccgcacagcttacccaaatgccttтc 502
             ||     |       |   ||  |  ||| |   | |||| ||  ||||||   |  |
Db     353 atcgatctgtgagattttattacatcgtagaaagtagaacagtttggccaaatttatggc 412

Qy     503 gcattgccaagctgatgctttacatttttgtcgtcatccattggaacagctgcctatact 562
           |  |||  || || || | |  | | ||  |  |    || |||  | | ||  | || |
Db     413 gtgttgttaacctaattcatatcctgttaatactggcacactggttcggttgtttctatt 472

Qy     563 ttgccctatcccggtacctgggcttcgggcgtgacgcatgggtgtacccggaccccgcgc 622
           ||   || |||  |     ||||||   | | |||   ||||||| ||| | |   |
Db     473 ttttactctccgaggcggagggctttcagggcgac---tgggtgtatccgtatcgacccg 529

Qy     623 agcctggctttgagcgcctgcggcgccagtacctctatagcttttacttctccacgctga 682
             |  | || ||    |||  |||  |||| ||    ||| | ||||  || || ||||
Db     530 gg---gactatgccaccctaacgcgaaagtatctgggcagcctgtactggtcgaccctga 586

Qy     683 tactgactacagtgggcgatacaccgccgccagccagggaagaagagtacctcttcatgg 742
            ||||||  ||  | ||  ||    ||  | || |  | | | | | ||| ||  | || | |
Db     587 cactgacgaccattggggacctgcccacacccgaaacgaatgcagaatatatttttacga 646

Qy     743 tgggcgacttcctgctggccgtcatgggtttcgccaccatcatgggtagcatgagctctg 802
           | |    ||   || | |  ||  |    ||||||||||  ||||     || ||  ||
Db     647 tcgttagctatttgattggtgtttttatcttcgccaccattgtgggccaagtgggcaatg 706

Qy     803 tcatctacaacatgaacactgcagatgcggctttctacccagatcatgcactggtgaaga 862
           | ||    |||   ||  |          ||  ||| | |   ||  |    ||  ||||
Db     707 tgataacgaaccgaaatgcgaatcgcctggagttcgagcgccttctggatggggccaaga 766

Qy     863 agtacatgaagctgcagcacgtcaaccgcaagctggagcggcgagttattgactggtatc 922
            || |||  ||  ||  | ||   | |   | || |||||||  ||  |    |||||

Db     767 cctatatgcggcaccacaaggtgcccggagggatgaagcggcgcgtgctgcgatggtacg 826

Qy     923 agcacctgcagatcaacaagaagatgaccaacgaggtagccatc---------ttacagc 973
           |  |       |  |  |   | |||       |  |  | ||||         |    |
Db     827 actatagctggtcccgcggcaggatacagggtggcggtgacatcaataccgcactgggcc 886

Qy     974 acttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctgtccactctgagcc 1033
             | |||| ||   ||||    | ||| ||||  |   |||  |||||  |  |||||
Db     887 tcctgcccgacaagctgaaaaccgaattggccttacatgtcaacctgagcgttctgaaga 946

Qy    1034 gggtgcagatctttcagaactgtgaggccagcctgctggaggagctggtgctgaagctgc 1093
           ||||   || || ||  | ||  || ||    | ||  | || || ||||| ||| ||
Db     947 aggtgaccatattccaagagtgccagcccgagttccttcacgatctcgtgctcaagatga 1006

Qy    1094 agccccagacctactcaccaggtgaatatgtatgccgcaaaggagacattggccaagaga 1153
           || || | | ||   | || || || |   | ||||| || || ||  | | ||  ||||
Db    1007 aggcctacatctttacgccgggcgactccatttgccgtaagggcgaggtggcccgcgaga 1066

Qy    1154 tgtacatcatccgagagggtcaactggccgtggtggcagatgatggtatcacacagtatg 1213
           ||| ||||||    || ||    |||| ||| ||   ||    || |     |
Db    1067 tgttcatcattgccgatggcattctggaggtgctgagcgagacgggcaaggtcctaacca 1126

Qy    1214 ctgtgctcggtgcagggctctactttggggagatcagcatcatcaacatcaaagggaaca 1273
           |  ||      || ||    |  ||||| || ||| ||||| |||| |  | |||
Db    1127 ccatg---aaggctggcgatttttttggcgaaatcggcatcctcaatctggacggg---- 1179

Qy    1274 tgtctgggaaccgccgcacagccaacatcaagagcctaggttattcagacctattctgcc 1333
                 |||   ||||||||  |  |      | | || || || || || |||| |
Db    1180 -----cttaacaaacgcacagcggatgtgcgctccgtgggctactcggagctcttctctc 1234

Qy    1334 tgagcaaggaggacctgcgggaggtgctgagcgagtatccacaagcacagaccatcatgg 1393
           ||  |   |||||  | | || ||   |||  || || ||  | || |||   ||  ||
Db    1235 tgtcccgcgaggatgtcctggcggccatgaaggactaccccgatgcccaggagattctgc 1294

Qy    1394 aggagaaaggacgtgagatcctgctgaaaatgaacaagttggacgtgaatgctgaggcag 1453
           ||      || ||  ||   ||  || |  ||       || ||   ||    |    |
Db    1295 agaccctgggccgcaagcggctaatggaggtgcgttgcgtgaacaaaaagtacgccaagg 1354

Qy    1454 ctgagatcgccctgcaggaggccac 1478
           |  |||  | |  | ||| |||| |
Db    1355 cgcagagtgacaaggaggcggccgc 1379


RESULT   8
ABL13352
ID   ABL13352 standard; cDNA; 8585 BP.
XX
AC   ABL13352;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 34538.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB69249.
```

```
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 34538; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 8585 BP; 2404 A; 1971 C; 1922 G; 2288 T; 0 other;


  Query Match             11.1%;  Score 191.2;  DB 23;  Length 8585;
  Best Local Similarity   55.2%;  Pred. No. 1.3e-40;
  Matches  373;  Conservative    0;  Mismatches  303;  Indels    0;  Gaps    0;

Qy    787 ggtagcatgagctctgtcatctacaacatgaacactgcagatgcggctttctacccagat 846
          |||| |||  ||||  | |||| |||||||||    ||    | ||  ||| |
Db   5569 ggtaacattggctccatgatctccaacatgaatgtggcccgcgtggagttccagaatcgc 5628

Qy    847 catgcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggagcggcga 906
            |    || ||| ||||||||    | | ||  ||    | | ||||||||   ||
Db   5629 atggacggagtcaagcagtacatggccttccggcgagtgggtcacgagctggaggcccgg 5688

Qy    907 gttattgactggtatcagcacctgcagatcaacaagaagatgaccaacgaggtagccatc 966
          || ||   ||||| |    ||  |  |    | |       |    ||||||      |
Db   5689 gtgatccgctggtttgcctacacgtggtcgcagagtggtgcgctggacgaggagcgtgtg 5748

Qy    967 ttacagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctgtccact 1026
           |        |||| ||   ||||  ||| ||  | ||| |    |||||| ||   ||
Db   5749 ctggccgctctgccggacaagctgaaggcggagatcgctatccaggtgcacatggatacg 5808

Qy   1027 ctgagccgggtgcagatctttcagaactgtgaggccagcctgctggaggagctggtgctg 1086
          ||||  | |||||  ||||| ||  ||   ||  |  | || || ||||  |||||||||
Db   5809 ctgaagcaggtgcgcatcttccacgacacggaaccgggtctcctcgaggccctggtgctg 5868

Qy   1087 aagctgcagccccagacctactcaccaggtgaatatgtatgccgcaaaggagacattggc 1146
          |||||  | |  |||  || |   || || || ||| | || ||||| || ||| | |||
Db   5869 aagctcaaactgcaggtcttcagtccgggcgactatatctgtcgcaagggtgacgtgggc 5928

Qy   1147 caagagatgtacatcatccgagagggtcaactggccgtggtggcagatgatggtatcaca 1206
           | ||||||||||| |       ||  |  || |||| || |  |||||||| |||||
Db   5929 aaggagatgtacatcgtgaagcgtggcaagttgtccgtcgtcggtgatgatggcatcacc 5988

Qy   1207 cagtatgctgtgctcggtgcagggctctactttggggagatcagcatcatcaacatcaaa 1266
            |   ||     | || || ||       || || ||| |   | |  || | ||
Db   5989 gtgctggccaccttgggcgccggttcggtgttcggcgaggtttccgtactcgagattgcg 6048

Qy   1267 gggaacatgtctgggaaccgccgcacagccaacatcaagagcctaggttattcagaccta 1326
          || ||   | | || || || ||||| |||||  | |    ||| || || || || ||
Db   6049 ggcaatcggacgggcaatcggcgcaccgccaatgtgagatccctgggctactcggatctc 6108

Qy   1327 ttctgcctgagcaaggaggacctgcgggaggtgctgagcgagtatccacaagcacagacc 1386
          |||||  |||   ||   |||  || |||||   |||  |||  |||||  | ||     ||
Db   6109 ttctgtctggcgaaacgggatttgtgggagaccctgtccgactatccggaggccagatcc 6168

Qy   1387 atcatggaggagaaaggacgtgagatcctgctgaaaatgaacaagttggacgtgaatgct 1446
          |   |     ||  |||| |  || | |||| ||      |    |||||| | |
Db   6169 acgctcactcagcgaggatgccagctgctgcgcaaggatggcttactggacgagcaaata 6228

Qy   1447 gaggcagctgagatcg 1462
             || | | || |||
Db   6229 ttcgctggtaagttcg 6244
```

```
RESULT   9
AAS91258
ID   AAS91258 standard; cDNA; 1125 BP.
XX
AC   AAS91258;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #27062.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG27071.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 27062; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1125 BP; 283 A; 276 C; 228 G; 338 T; 0 other;


  Query Match              9.8%;  Score 169.2;  DB 23;  Length 1125;
  Best Local Similarity   55.0%;  Pred. No. 4.2e-35;
  Matches  333;  Conservative    0;  Mismatches  273;  Indels    0;  Gaps    0;

Qy    272 gattcttggaacagggcatcctggtggtggacaagggtaggatctcgagtcgctacgttc 331
          | | |||  | || ||  |  ||||   | |  |   ||   ||   | |   ||
Db    515 gttacttaaaccaaggactgttggtaaagaaagaacttaaactcataaataaatataaat 574
```

```
Qy    332 gcacctggagtttcttcttggacctggcttccctgatgcccacagatgtggtctacgtgc 391
          || || |   ||     | ||  |    || ||||| || || ||| || | ||  |
Db    575 ccaacttgcaatttaaacttgatgttctgtcactgataccaactgatttgctgtatttta 634

Qy    392 ggctgggcccgcacacacccaccctgaggctgaaccgctttctccgcgcgccccgcctct 451
          | | ||   | ||   ||     | ||  | ||| | ||  | ||     | ||  | |
Db    635 agttagggtggaactatccagaaattagattaaacaggttgttacggttctctcgtatgt 694

Qy    452 tcgaggccttcgaccgcacagagacccgcacagcttacccaaatgcctttcgcattgcca 511
          | |||  |||| |  | ||||| ||  | |||   || |||||   |||  | ||| |||
Db    695 ttgagttcttccagagaacagaaacaaggacaaactatccaaacatcttcaggatttcca 754

Qy    512 agctgatgctttacatttttgtcgtcatccattggaacagctgcctatactttgccctat 571
          | ||  |  | || ||  |  || | ||||| |||||    ||  | | ||   |  | |
Db    755 accttgttatgtatatcgtcatcattatccactggaatgcatgtgtgttctactctattt 814

Qy    572 cccggtacctgggcttcgggcgtgacgcatgggtgtacccggaccccgcgcagcctggct 631
          |        | || || ||   ||| |||||||| ||||| ||        | ||||  |
Db    815 ctaaagctattggatttggaaatgatacatgggtctaccctgatattaatgatcctgaat 874

Qy    632 ttgagcgcctgcggcgccagtacctctatagcttttacttctccacgctgatactgacta 691
          |||  ||  ||    |  | ||| | || ||| ||||||  || || ||||   ||||||
Db    875 ttggccgtttggctagaaaatacgtatacagcctttactggtctacactgactttgacta 934

Qy    692 cagtgggcgatacaccgccgccagccagggaagaagagtacctcttcatggtgggcgact 751
          |  | || || ||||| || || |  |||||    |||||  ||||  ||||||  || |
Db    935 ccattggtgaaacacccctcccgtgagggattctgagtatgtctttgtggtggttgatt 994

Qy    752 tcctgctggccgtcatgggtttcgccaccatcatgggtagcatgagctctgtcatctaca 811
          ||||  | |  ||  |   ||| || |||||| | |||| |||  | ||| | || | ||
Db    995 tcctaattggagtgttaatttttgctaccatcgttggtaacataggttctatgatttcca 1054

Qy    812 acatgaacactgcagatgcggctttctacccagatcatgcactggtgaagaagtacatga 871
          |||||||  | ||    || |  ||  |  ||     ||       | ||| | || |||
Db   1055 acatgaatgcagccagagcagaatttcaagcaagaattgatgctatcaagcaatatatgc 1114

Qy    872 agctgc 877
          |  | |
Db   1115 attttc 1120


RESULT  10
ABL06311
ID   ABL06311 standard; cDNA; 1281 BP.
XX
AC   ABL06311;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 13415.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB62208.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 13415; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1281 BP; 315 A; 345 C; 384 G; 237 T; 0 other;


  Query Match             9.5%;  Score 163.6;  DB 23;  Length 1281;
  Best Local Similarity   56.6%;  Pred. No. 1.4e-33;
  Matches  326;  Conservative    0;  Mismatches  244;  Indels    6;  Gaps    1;

Qy    856 gtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggagcggcgagttattgac 915
          ||||||| |||||||  | |||  |  ||         | ||| ||  |   || ||  |
Db     13 gtgaagacgtacatgcggatgcgacgtgtgccgaatcatctgcaggtgaaggtcatcaaa 72

Qy    916 tggtatcagcacctgcagatcaacaagaagatgaccaacgaggtagccatcttacagcac 975
          ||||   |  |||||  | | |   | ||     |  ||||||    |  | |
Db     73 tggttcgattacctgtggcttacgcaaaaatgctcggacgaggagcgcgccgtgtcctgt 132

Qy    976 ttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctgtccactctgagccgg 1035
          | ||||||     |   ||| ||| | ||  |    || ||  |    || || |  |||
Db    133 cttcctgataaattaaaggctgaaatagcaattaacgtccatttagatacgctcaagcgg 192

Qy   1036 gtgcagatctttcagaactgtgaggccagcctgctggaggagctggtgctgaagctgcag 1095
          ||| |||| || || |||   |||||| || | ||    || |||||||||   ||| |
Db    193 gtggagattttccaaaacacagaggccggcttcctatgcgaactggtgctgcgcctgagg 252

Qy   1096 ccccagacctactcaccaggtgaatatgtatgccgcaaaggagacattggccaagagatg 1155
          |||  |  || ||| || || || ||  | ||| | || || ||  | ||| | ||||||
Db    253 cccgtgctcttctcgcccggcgactacatctgcagaaagggcgaggtgggcaaggagatg 312

Qy   1156 tacatcatccgagagggtcaactggccgtggtggcagatgatggtatcacacagtatgct 1215
          |||||  |       || | | ||   |||||||| ||  |||| |  ||   |   ||
Db    313 tacattgtgaaccgaggacgattgcaggtggtggccgacaatggaaagacggtgatggcc 372

Qy   1216 gtgctcggtgcagggctctactttggggagatcagcatcatcaacatcaaagggaacatg 1275
             ||    || ||   ||| ||||| ||||| || ||  |||| ||      |  ||
Db    373 tccctgaaggctggttcctattttggcgagattagtatactcaatat------gggcacc 426

Qy   1276 tctgggaaccgccgcacagccaacatcaagagcctaggttattcagacctattctgcctg 1335
          | || ||||| ||||||||||| | |        | || ||    |||||  |||  ||||
Db    427 gcaggcaaccgacgcacagccagcgttcgctcagtgggatacagcgacctcttcgtcctg 486

Qy   1336 agcaaggaggacctgcgggaggtgctgagcgagtatccacaagcacagaccatcatggag 1395
          |||||| ||||| || |||| |||||||  ||||||||    || |         |||||
Db    487 agcaagaaggacatgtgggacgtgctgaaggagtatccggcggcgcgtgttcgtctggag 546

Qy   1396 gagaaaggacgtgagatcctgctgaaaatgaacaag 1431
             || ||      ||    ||  ||||   || |||
Db    547 tcgatagccgtcaagcgattggagaaatacaagaag 582


RESULT  11
ABL18084/c
ID   ABL18084 standard; DNA; 17150 BP.
XX
AC   ABL18084;
XX
DT   26-MAR-2002  (first entry)
XX
```

```
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 5725.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 5725; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 17150 BP; 4627 A; 3695 C; 3920 G; 4908 T; 0 other;
```

Query Match 8.2%; Score 141; DB 23; Length 17150;
Best Local Similarity 51.3%; Pred. No. 4.9e-27;
Matches 386; Conservative 0; Mismatches 355; Indels 12; Gaps 2;

```
Qy      43 ccagccccatccaaggccaggaagttgctgcctgtcctggacccatctggggattactac 102
           | |||  || |  |||||||||  |||     ||| | || ||  | |||      | |
Db   16222 CGAGCAGCAACGGCGGCCAGGAAACTGCACTTCGTCTTTGATCCGGCGGGGCGACTGTGC 16163

Qy     103 tactggtggctgaacacaatggtcttcccagtcatgtataacctcatcatcctcgtgtgc 162
           ||||  |||   |      ||  | |  |  || |||| ||| ||    |  |  | | |
Db   16162 TACTACTGGTCCATGGTGGTGTCCATGGCCTTCCTGTACAACTTCTGGGTGATAATCTAC 16103

Qy     163 agagcctgcttccccgacttgcagcacggttatctggtggcctggttggtgctggactac 222
           |   |  |||||  ||  |  | |   |     | |     |||||    |||||||||
Db   16102 CGCTTCGCCTTCCAGGAGATCAATCGGCGAACGATCGCCATTTGGTTCTGTCTGGACTAC 16043

Qy     223 acgagtgacctgctatacctactagacatggtggtgcgcttccacacaggattcttggaa 282
             |  ||| | || || ||  | || ||  || | | ||| | ||| |||| | ||||
Db   16042 CTGTCCGACTTCCTGTATCTGATCGATATACTGTTCCACTTTCGCACCGGATACCTGGAG 15983

Qy     283 cagggcatcctggtggtggacaagggtaggatctcgagtcgctacgttcgcacctggagt 342
           | ||  | |||  |  ||||     | | |    |  | |||| |   | |   ||
Db   15982 GACGGGGTGCTGCAGACGGACGCTTTGAAGCTGCGCACCCACTACATGAACTCGACGATC 15923

Qy     343 ttcttcttggacctggcttccctgatgcccacagatgtggtctacgtgcggctgggcccg 402
           |||| | | |||      | |||| ||||    ||  |  ||||  ||    | |||
Db   15922 TTCTACATCGACTGCCTGTGCCTGCTGCCGCTGGACTTTCTCTATTTGTCCATCGGCTTT 15863

Qy     403 cacacacccaccctgaggctgaaccgctttctccgcgcgccccgcctcttcgaggccttc 462
             ||    | | |||| |      |||  |  |         |||  |||  |    | |
Db   15862 AAC---TCGATCCTGCGCAGCTTCCGGCTGGTGAAGATCTACCGGTTCTGGGCCTTCATG 15806

Qy     463 gaccgcacagagacccgcacagcttacccaaatgcctttcgcattgccaagctgatgctt 522
           || ||||| |||   | |||    ||||| ||    || || |   |    || || |
Db   15805 GATCGCACCGAGCGGCACACCAACTACCCGAACCTGTTCCGGAGCACGGCCCTCATCCAC 15746

Qy     523 tacatttttgtcgtcatccattggaacagctgcctatactttgccctatcccggtacctg 582
           ||| |  |||| |  ||||||||||| | || || |||     | ||  |  | ||
Db   15745 TACCTGCTTGTGATATTCCATTGGAACGGTTGTCTCTACCACATCATACACAAGAACAAT 15686

Qy     583 ggcttcgggcgtgacgcatgggtgtacccggaccccgcgcagcctggctttgagcgcctg 642
           || |||||       |   ||||| |||  ||| | | |  | | | | | |
Db   15685 GGGTTCGGATCACGCAACTGGGTCTACCACGACTCGGAGTCGGCGGACGTGGTCAAG--- 15627

Qy     643 cggcgccagtacctctatagcttttacttctccacgctgatactgactacagtgggcgat 702
                 ||||| ||  | ||||  ||||  | ||| |||   ||||| ||  | || |||
Db   15628 ------CAGTATCTGCAGAGCTACTACTGGTGCACCCTGGCCCTGACCACCATCGGGGAT 15575

Qy     703 acaccgccgccagccagggaagaagagtacctcttcatggtgggcgacttcctgctggcc 762
             ||   |||   |    | |  |||||| |||| |  |     |  | ||  | | |
Db   15574 CTTCCCAAGCCGCGCTCCAAGGGCGAGTACGTCTTTGTTATCCTGCAGCTGCTCTTTGGC 15515

Qy     763 gtcatgggtttcgccaccatcatgggtagcatg 795
           | |||   |||||||  |  ||||   | ||
Db   15514 CTGATGCTCTTCGCCACGGTACTGGGGCACGTG 15482
```

```
RESULT  12
ABL19457
ID   ABL19457 standard; DNA; 3275 BP.
XX
AC   ABL19457;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 9844.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 9844; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
```

```
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3275 BP; 843 A; 849 C; 873 G; 710 T; 0 other;

  Query Match             7.9%;  Score 137;  DB 23;  Length 3275;
  Best Local Similarity   45.7%;  Pred. No. 2.8e-26;
  Matches  626;  Conservative    0;  Mismatches  710;  Indels   33;  Gaps    3;

Qy      58 gccaggaagttgctgcctgtcctggacccatctggggattactactactggtggctgaac 117
           ||||  ||      ||    |||||| ||    ||  |   |||| ||  |||||   |
Db    1512 gccaacaatgatgagcgaaccctggatccccagggcaagatctacatctcctggctatgc 1571

Qy     118 acaatggtcttcccagtcatgtataacctcatcatcctcgtgtgcagagcctgcttcccc 177
           |     |  |  |  | ||||||  |    |  ||  |    | ||||  || |||
Db    1572 gttgtatcgctatcgtttctctataacgcctgggtaatcccgctgcgtgcctcgtttccc 1631

Qy     178 gacttgcagcacggttatctggtggcctggttggtgctggactacacgagtgacctgcta 237
                | |   |         |||| |||     ||||       |||| |  ||
Db    1632 tttcaaaccaaggagaacaccaatatctggctggcctgcgacttttgcgctgacatcata 1691

Qy     238 tacctactagacatggtggtgcgcttccacacag---gattcttggaacagggcatcctg 294
           || ||||| ||  |||| |       ||    |     | | |     ||||  ||  |
Db    1692 tatctactggatgtggtcttctttaaacatcgggttatgtacctttttgagggattctgg 1751

Qy     295 gtggtggacaagggtaggatctcgagtcgctacgttcgcacctggagtttcttcttggac 354
           |||    |||||      | |   |    |||| | || |    |   ||     ||||
Db    1752 gtgaaaaacaagaacctcaccagaaagaactacatgcgaaaactgcagtttaagctggat 1811

Qy     355 ctggcttccctgatgcccacagatgtggtctacgtgcggctgggcccgcacacacccacc 414
           |||    ||||  | ||    ||  |  | ||| |    || ||  | ||  |
Db    1812 ctgctggccctttcttccgctggagctattgtactttaaactcggaacacaggcggtttgg 1871

Qy     415 ctgaggctgaaccgctttctccgcgcgccccgcctcttcgaggccttcgaccgcacagag 474
           || |  |     | ||  |       |   || |||  || |  ||   ||     ||
Db    1872 ttgcgttttccaaggttctttaagattcagagcttctgggaagtgtttcgccttttggat 1931

Qy     475 acccgcacagcttacccaaatgcctttcgcattgccaagctgatgctttacatttttgtc 534
              || | |||  ||  |     ||||  | |||||     |    |||||  |   |
Db    1932 cgagtcatatcttctccgcactttgttcgagtggccaaaaccctaacctacatgctctac 1991

Qy     535 gtcatccattggaacagctgcctatactttgccctatcccggtacctgggcttcgggcgt 594
            | |||||    |        || ||||  |||         |||| |||  | || |
Db    1992 atgatccacatcacggctgctctgtactacgcctatagtgactaccagggtctgggacag 2051

Qy     595 gacgcatgggtgtacccggaccccgcgcagcctggctttgagcgcctgcggcgccagtac 654
            |    ||||| | |   | |   | ||| ||                        ||
Db    2052 aatcgttgggtcttcagcggcaaagggcatcc------------------------ctat 2087

Qy     655 ctctatagcttttacttctccacgctgatactgactacagtgggcgatacaccgccgcca 714
            |     |||||   ||  ||||   ||     ||| |  | ||  | |  ||   |||
Db    2088 gtaaggtgctttgcttttgccaccaagacggcaacttcgataggaaaaaaccccaagccg 2147

Qy     715 gccagggaagaagagtacctcttcatggtgggcgacttcctgctggccgtcatgggtttc 774
           |   |  | | |||||||| ||||||||  ||  | ||  ||  ||| |||| | |  |||
Db    2148 gaacgtcagggagagtacgtcttcatgacggtggcctggctaatgggcgtctttgtgttc 2207

Qy     775 gccaccatcatgggtagcatgagctctgtcatctacaacatgaacactgcagatgcggct 834
           ||     | || |||   || ||   | ||||||  |     |         |    |
Db    2208 gctctgcttatcggtcagatcagggatatcatctcgacagccactcggaataagcatgag 2267

Qy     835 ttctacccagatcatgcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaag 894
           |    ||      | |    |     | ||||||||   |  |   |  ||   |   ||
Db    2268 tatcgccagctggaggatgagacactggagtacatgcgacgcctcaatctctcgcaggag 2327

Qy     895 ctggagcggcgagttattgactggtatcagcacctgcagatcaacaagaagatgaccaac 954
            || ||  | | || |     ||||  |||        |    |  ||   |      |
Db    2328 gtgcagtcgagggtcaaaatgtggttccagtttacatgggagcagcagcgcactttggat 2387

Qy     955 gaggtagccatcttacagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtg 1014
           ||      ||| ||  |    |||||         ||    | ||  | ||  | || |||
Db    2388 gaatccaacatattggatgctttgcccatcaatttgaaaaccgatatcgccatctccgtg 2447

Qy    1015 cacctgtccactctgagccgggtgcagatctttcagaactgtgaggccagcctgctggag 1074
           ||| |    || ||   |  ||||||  |||||    | || ||||       | ||
Db    2448 cacatccaaacgctctccaaggtgcaactctttgctgattgcgaggaggctttactcaga 2507

Qy    1075 gagctggtgctgaagctgcagccccagacctactcaccaggtgaatatgtatgccgcaaa 1134
           || ||||| || || ||   | |    || |  |  ||||| || |  || |||||||||
Db    2508 gatctggttctaaaactaagggcagtcacatttttgccaggagatttcgtttgccgcaaa 2567

Qy    1135 ggagacattggccaagagatgtacatcatccgagagggtcaactggccgtggtggcagat 1194
           || ||  | |||  |||||||||| ||| |    |  |||  ||  |    ||| |||  | |
Db    2568 ggcgaagtgggcagagagatgtatatcgtgaaactgggacaggttcaagtgatgggtggt 2627

Qy    1195 gatggtatcacacagtatgctgtgctcggtgcagggctctactttggggagatcagcatc 1254
              | ||| |       |||  ||||| |   |      | |   |   | | || |
Db    2628 cctagtagcg------atgtagtgctggccactctaaccgaaggttccgtgttcggggag 2681

Qy    1255 atcaacatcaaagggaacatgtctgggaaccgccgcacagccaacatcaagagcctaggt 1314
           ||||   |    || |  |     |   | ||| | |||||  |  ||     |   ||
Db    2682 atcagtctgctgggaataaatggagccgatcgcagaacagcagatgtccgatccaagggc 2741

Qy    1315 tattcagacctattctgcctgagcaaggaggacctgcgggaggtgctgagcgagtatcca 1374
           || ||  |||| |||   |||  |||   |||  |    |||||  |   | | |||||
Db    2742 tactccaacctgttcgtgctgtccaaatcggatttaaacgaggtcattgcctactatccc 2801

Qy    1375 caagcacagaccatcatggaggagaaaggacgtgagatcctgctgaaaa 1423
              || ||| | ||  |  | ||   |  ||| || |  |||  ||||
Db    2802 acggcccaggcgattctcaaaaagcgtgctcgtcagctaatgcgaaaaa 2850


RESULT  13
AAS91658
ID   AAS91658 standard; cDNA; 3811 BP.
XX
AC   AAS91658;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #27462.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG27471.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 27462; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
```

```
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3811 BP; 812 A; 1093 C; 1102 G; 804 T; 0 other;


  Query Match             7.2%;  Score 124.8;  DB 23;  Length 3811;
  Best Local Similarity   46.4%;  Pred. No. 5.5e-23;
  Matches  586;  Conservative    0;  Mismatches  642;  Indels   36;  Gaps    4;

Qy    171 cttccccgacttgcagcacggttatctggtggcctggttggtgctggactacacgagtga 230
          ||||||| ||  |   |  |   |  |      |||| || || |||| |||    | ||
Db   2092 cttcccctaccagaccccggacaacatccaccactggctgctgatggattacctatgcga 2151

Qy    231 cctgctatacctactagacat---ggtggtgcgcttccacacaggattcttggaacaggg 287
          |||  | ||| | || |||||    ||| | |     | |      ||  |   |   ||
Db   2152 cctcatctacttcctggacatcaccgtgttccagacacgcctgcagtttgtcagaggcgg 2211

Qy    288 catcctggtggtggacaagggtaggatctcgagtcgctacgttcgcacctggagtttctt 347
            | |  |  ||||||       ||    | |  |||| |     |  |  | |||
Db   2212 ggacatcattacggacaaaaaggacatgcgaaataactacctgaagtctcgccgcttcaa 2271

Qy    348 cttggacctggcttccctgatgcccacagatgtggtctacgtgcggctgggcccgcacac 407
            ||||||||     |||  |||||   ||| |  ||||  ||    | ||   | ||
Db   2272 gatggacctgctcagcctcctgcccttggattttctctatttgaaagtcggtgtgaac-- 2329

Qy    408 acccaccctgaggctgaaccgctttctccgcgcgccccgcctcttcgaggccttcgaccg 467
           |||  |||  | |||  ||||| | |           | ||||||||     |  |||
Db   2330 -ccctcctccgcctgccccgctgtttaaagtacatggccttcttcgagtttaacagccg 2388

Qy    468 cacagagacccgcacagcttacccaaatgcctttcgcattgccaagctgatgctttacat 527
          |   ||  ||  |       |  |  | |  |   |  |   || |  '|    '||| |
Db   2389 cctggaatccatcctcagcaaagcctacgtgtacagggtcatcaggaccacagcctacct 2448

Qy    528 ttttgtcgtcatccattggaacagctgcctatactttgccctatcccggtacctgggctt 587
          | |   |  | | |||| |||   ||| || || |       |||    || | |||| |
Db   2449 tctctacagcctgcatttgaattcctgtctttattactgggcatcggcctatcagggcct 2508

Qy    588 cgggcgtgacgcatgggtgtacccggaccccgcgcagcctggctttgagcgcctgcggcg 647
          |||                          | || | ||    |  |  ||  || || |
Db   2509 cgg-----------------------ctccactcactgggtttacgatggcgtgggaaa 2544

Qy    648 ccagtacctctatagcttttacttctccacgctgatactgactacagtgggcgatacacc 707
          |   ||  |     | |  |||||  |   |  ||  || |  ||  | || |     ||
Db   2545 cagttatattcgctgttactactttgctgtgaagaccctcatcaccatcgggggctgcc 2604

Qy    708 gccgccagccagggaagaagagtacctcttcatggtgggcgacttcctgctggccgtcat 767
            ||    |       ||     |||||  | ||    | |   |   || |||| |
Db   2605 tgaccccaagacactctttgaaattgtcttccagctgctgaattatttcacgggcgtctt 2664

Qy    768 gggtttcgccaccatcatgggtagcatgagctctgtcatctacaacatgaacactgcaga 827
           | |||| |    || || ||    |||||   |||  |     |   | | | | | |
Db   2665 tgctttctctgtgatgatcggacagatgagagatgtggtaggggccgccaccgcgggaca 2724

Qy    828 tgcggctttctacccagatcatgcactggtgaagaagtacatgaagctgcagcacgtcaa 887
            |    | |  |         |       |  ||||||||||||  |  |  |  ||
Db   2725 gacctactaccgcagctgcatggacagcacggtgaagtacatgaatttctacaagatccc 2784

Qy    888 ccgcaagctggagcggcgagttattgactggtatcagcacctgcagatcaacaagaagat 947
          |       || ||   || || |    ||||||  || ||     |  |    |    ||

Db   2785 caagtccgtgcagaaccgcgtcaagacctggtacgagtacacctggcactcgcaaggcat 2844

Qy    948 gaccaacgaggtagccatcttacagcacttgcctgagcggctgcgggcagaagtggctgt 1007
          |    | |||  ||   |  |   |||  | || ||   | |||||   ||  | ||  |
Db   2845 gctggatgagtcagagctgatggtgcagcttccagacaagatgcggctggacctcgccat 2904

Qy   1008 gtctgtgcacctgtccactctgagccgggtgcagatctttcagaactgtgaggccagcct 1067
             ||| ||     ||   | |||   ||     ||||||||  ||||||         |
Db   2905 cgacgtgaactacaacatcgttagcaaagtcgcactctttcagggctgtgaccggcagat 2964

Qy   1068 gctggaggagctggtgctgaagctgcagccccagacctactcaccaggtgaatatgtatg 1127
          | |    ||  || ||  || ||| |   |      ||||   ||    || ||||| ||
Db   2965 gatctttgacatgctgaagaggcttcgctctgttgtctacctgcccaacgactatgtgtg 3024

Qy   1128 ccgcaaaggagacattggccaagagatgtacatcatccgagagggtcaactggccgtggt 1187
          |   || || || || ||||  ||||||||||||||||  |  || ||| ||   || |
Db   3025 caagaaggggggagatcggccgtgagatgtacatcatccaggcagggcaagtgcaggtctt 3084

Qy   1188 ggcagatgatggtatcacacagtatgctgtgctcggtgcagggctctactttggggagat 1247
          ||  |    || |   | |       |||||||  |   |   |           | | |
Db   3085 gggcggccctgatgggaaat------ctgtgctggtgacgctgaaagctggatctgtgtt 3138

Qy   1248 cagcatcatcaacatcaaagggaacatgtctgggaaccgccgcacagccaacatcaagag 1307
            |   || | | |    |       |   |||||||| |||| |||||| |   |
Db   3139 tggagaaataagcttgctggctgttgggggcgggaaccggcgcacggccaacgtggtggc 3198

Qy   1308 cctaggttattcagacctattctgcctgagcaaggaggacctgcgggaggtgctgagcga 1367
          |  || | | |  |||| |||  ||||   ||| |||||||   ||| |  ||    |
Db   3199 gcacgggtttaccaacctcttcatcctggataagaaggacctgaatgagattttggtgca 3258

Qy   1368 gtatccacaagcacagaccatcatggaggagaaaggacgtgagatcctgctgaaaatgaa 1427
           |||||  |  | ||||   |  |  | ||||||   |    || |||  |  |  ||
Db   3259 ttatcctgagtctcagaagttactccggaagaaagccaggcgcatgctgagaagcaacaa 3318

Qy   1428 caag 1431
           |||
Db   3319 taag 3322


RESULT  14
AAS79945
ID   AAS79945 standard; cDNA; 1146 BP.
XX
AC   AAS79945;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #15749.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG15758.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
```

PT biodiversity -
XX
PS Claim 1; SEQ ID No 15749; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1146 BP; 357 A; 219 C; 266 G; 304 T; 0 other;

Query Match 6.9%; Score 118.4; DB 23; Length 1146;
Best Local Similarity 70.5%; Pred. No. 1.7e-21;
Matches 158; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Qy 1104 ctactcaccaggtgaatatgtatgccgcaaaggagacattggccaagagatgtacatcat 1163
Db 360 ctacagtcctggagattatatttgcaagaaaggggatatcggacgagagatgtacattat 419

Qy 1164 ccgagagggtcaactggccgtggtggcagatgatggtatcacacagtatgctgtgctcgg 1223
Db 420 caaggaaggcaaactcgctgtggtggcagatgatggagtcactcagtttgtggtattgag 479

Qy 1224 tgcagggctctactttggggagatcagcatcatcaacatcaaagggaacatgtctgggaa 1283
Db 480 cgatggcagctacttcggtgagatcagcattcttaacattaaagggagcaaagctggcaa 539

Qy 1284 ccgccgcacagccaacatcaagagcctaggttattcagacctat 1327
Db 540 tcgaagaacggccaatattaaaagtattggctactcagacctgt 583

RESULT 15
AAS77668
ID AAS77668 standard; cDNA; 2421 BP.
XX
AC AAS77668;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #13472.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG13481.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 13472; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2421 BP; 715 A; 553 C; 591 G; 562 T; 0 other;

Query Match 6.9%; Score 118.4; DB 23; Length 2421;
Best Local Similarity 70.5%; Pred. No. 2.3e-21;
Matches 158; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Qy 1104 ctactcaccaggtgaatatgtatgccgcaaaggagacattggccaagagatgtacatcat 1163
Db 360 ctacagtcctggagattatatttgcaagaaaggggatatcggacgagagatgtacattat 419

Qy 1164 ccgagagggtcaactggccgtggtggcagatgatggtatcacacagtatgctgtgctcgg 1223
Db 420 caaggaaggcaaactcgctgtggtggcagatgatggagtcactcagtttgtggtattgag 479

Qy 1224 tgcagggctctactttggggagatcagcatcatcaacatcaaagggaacatgtctgggaa 1283
Db 480 cgatggcagctacttcggtgagatcagcattcttaacattaaagggagcaaagctggcaa 539

Qy 1284 ccgccgcacagccaacatcaagagcctaggttattcagacctat 1327
Db 540 tcgaagaacggccaatattaaaagtattggctactcagacctgt 583

Search completed: June 20, 2002, 14:39:36
Job time: 13574 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2002, 14:28:13 ; Search time 155.16 Seconds
(without alignments)
2735.594 Million cell updates/sec

Title: US-09-927-267-3
Perfect score: 1728
Sequence: 1 atgagccaggacaccaaagt.....gacccccaggtccagagtga 1728

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	78.4	4.5	1307	4	US-09-172-422-3	Sequence 3, Appli
	2	59.8	3.5	3480	4	US-09-226-012-1	Sequence 1, Appli
	3	59.8	3.5	3950	4	US-09-226-012-3	Sequence 3, Appli
c	4	52.8	3.1	7218	1	US-08-232-463-14	Sequence 14, Appl
	5	52.4	3.0	7218	1	US-08-232-463-14	Sequence 14, Appl
	6	52	3.0	3323	4	US-09-600-776-1	Sequence 1, Appli
	7	49.8	2.9	3064	4	US-09-600-776-5	Sequence 5, Appli
	8	49	2.8	3141	2	US-08-956-242-1	Sequence 1, Appli
	9	49	2.8	3141	3	US-09-351-215-1	Sequence 1, Appli
	10	44	2.5	4722	4	US-08-979-608A-14	Sequence 14, Appl
	11	41.8	2.4	809	2	US-08-370-909-1	Sequence 1, Appli
	12	41.8	2.4	809	4	US-09-056-105-22	Sequence 22, Appl
	13	41	2.4	1995	1	US-08-425-069-3	Sequence 3, Appli
	14	41	2.4	1995	2	US-08-317-844B-3	Sequence 3, Appli
	15	40.4	2.3	2580	3	US-09-050-863-2	Sequence 2, Appli
	16	40.4	2.3	2580	4	US-09-359-081-2	Sequence 2, Appli
c	17	40.4	2.3	5452	2	US-09-130-114-1	Sequence 1, Appli
	18	40.4	2.3	9600	4	US-08-910-647-1	Sequence 1, Appli
	19	40.4	2.3	10596	1	US-07-884-811-15	Sequence 15, Appl
	20	40.4	2.3	10596	1	US-07-885-971-15	Sequence 15, Appl
	21	40.4	2.3	10596	1	US-08-087-783A-15	Sequence 15, Appl
	22	40.4	2.3	10596	1	US-08-194-088B-15	Sequence 15, Appl
	23	40.4	2.3	10596	2	US-08-194-087-15	Sequence 15, Appl
	24	40.4	2.3	10596	5	PCT-US93-04648-15	Sequence 15, Appl
	25	39.8	2.3	5024	4	US-09-307-143-1	Sequence 1, Appli
	26	39	2.3	2338	1	US-08-425-069-1	Sequence 1, Appli
	27	39	2.3	2338	2	US-08-317-844B-1	Sequence 1, Appli
	28	38.6	2.2	1007	3	US-09-248-335-31	Sequence 31, Appl
	29	38.6	2.2	3736	4	US-09-600-776-10	Sequence 10, Appl
	30	38.2	2.2	2818	1	US-08-366-276-1	Sequence 1, Appli
	31	37.2	2.2	1599	1	US-08-288-405A-9	Sequence 9, Appli
	32	37.2	2.2	3177	1	US-08-042-747A-4	Sequence 4, Appli
	33	37.2	2.2	4403765	4	US-09-103-840A-2	Sequence 2, Appli
	34	36.8	2.1	2214	3	US-08-864-038A-1	Sequence 1, Appli
	35	36.8	2.1	3331	3	US-08-864-038A-2	Sequence 2, Appli
	36	36.8	2.1	3331	3	US-08-864-038A-4	Sequence 4, Appli
	37	35.8	2.1	1218	2	US-08-446-875-1	Sequence 1, Appli
	38	35.8	2.1	1218	2	US-08-102-385G-1	Sequence 1, Appli
	39	35.6	2.1	3142	2	US-08-956-242-3	Sequence 3, Appli
	40	35.6	2.1	3142	3	US-09-351-215-3	Sequence 3, Appli
	41	35.6	2.1	3635	2	US-08-588-983-15	Sequence 15, Appl
	42	35.6	2.1	3635	2	US-08-588-976-15	Sequence 15, Appl
	43	35.2	2.0	2057	3	US-09-059-522-2	Sequence 2, Appli
	44	35.2	2.0	2057	3	US-09-382-027-2	Sequence 2, Appli
	45	35.2	2.0	2385	1	US-08-393-333-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-172-422-3
; Sequence 3, Application US/09172422A
; Patent No. 6300485
; GENERAL INFORMATION:
; APPLICANT: Adams, Arwen E.
; APPLICANT: Chiu, Choi Ying
; APPLICANT: Duhl, David
; APPLICANT: Gorman, Susan W.
; APPLICANT: Leng, Song
; APPLICANT: Sheffield, Val
; APPLICANT: Welch, Juliet
; TITLE OF INVENTION: MYOSIN IXA AND CYCLIC NUCLEOTIDE GATED
; TITLE OF INVENTION: CHANNEL-15 (CNGC-15) POLYNUCLEOTIDES, POLYPEPTIDES,
; TITLE OF INVENTION: COMPOSITIONS, METHODS, AND USES THEREOF
; FILE REFERENCE: 200130.442
; CURRENT APPLICATION NUMBER: US/09/172,422A
; CURRENT FILING DATE: 1998-10-14
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1307
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1307)
; OTHER INFORMATION: n = A,T,C or G
US-09-172-422-3

Query Match 4.5%; Score 78.4; DB 4; Length 1307;
Best Local Similarity 46.6%; Pred. No. 5.1e-12;
Matches 324; Conservative 0; Mismatches 366; Indels 6; Gaps 2;

```
Qy    500 ttcgcattgccaagctgatgctttacatttttgtcgtcatccattggaacagctgcctat 559
          | ||||| |  || || ||       ||  |  |  ||  ||| ||| || |||||||
Db    590 tgcgcatcgtgaacctcatcggcatgatgctcctgctctgccactgggacggctgcctgc 649

Qy    560 actttgccctatcccggtacctgggcttcgggcgtgacgcatgggtgtacccggaccccg 619
          | ||     || ||  |   | || ||||      |||   ||||||| |    ||  |
Db    650 agttcctggtacccatgctacaggacttccctgacgactgctgggtgtccatcaacaaca 709

Qy    620 cgcagcctggctttgagcgcctgcggcgccagtacctctatagcttttacttctccacgc 679
          |  |     ||    | |   |  ||   | | |  |     |||  |   |   |  |
Db    710 tggtgaacaactcctgggg---gaagcagtactcctacgcgctcttcaaggccatgagcc 766

Qy    680 tgatactgactacagtgggcgatacaccgccgccagccagggaagaagagtacctcttca 739
            || |||   |  |    ||     | | |||| |   |         ||   |    |||
Db    767 acatgctgtgcatcggctacgggcggcaggcgcccgtgggcatctccgacgtctggctca 826
```

```
Qy    740 tggtgggcgacttcctgctggccgtcatgggtttcgccaccatcatgggtagcatgagct 799
          ||  |  | |  |  |||  | ||   | | |||||   |||| ||   |   |
Db    827 ccatgctcagcatgatcgtgggtgccacctgctacgccatgttcattggccacgccactg 886

Qy    800 ctgtcatctacaacatgaacactgcagatgcggctttctacccagatcatgcactggtga 859
          | ||||| |   | || || |  |          | | |   | |  |    | ||||
Db    887 ccctcatccagtccctggactcctcccggcgccagtaccaggaaaagtacaagcaggtgg 946

Qy    860 agaagtacatgaagctgcagcacgtcaaccgcaagctggagcggcgagttattgactggt 919
          || ||||||||    | ||  |  ||   | | |      || |||  |    ||||  |
Db    947 agcagtacatgtcctttcacaagctcccgcccgacacccggcagcgcatccacgactact 1006

Qy    920 atcagcacctgcagatcaacaagaagatgaccaacgaggtagccatcttacagcacttgc 979
          |  ||||||   |   | |    ||||||  | ||||||    || | |     |  |
Db   1007 acgagcaccgcta---ccagggcaagatgttcgacgaggagagcaccctgggcgagctaa 1063

Qy    980 ctgagcggctgcgggcagaagtggctgtgtctgtgcacctgtccactctgagccgggtgc 1039
          ||||  |||||||  ||  |       | |      | |       ||  |    |||
Db   1064 gcgagcccctgcgggaggagatcatcaactttaactgtcggaagctggtggcctccatgc 1123

Qy   1040 agatctttcagaactgtgaggccagcctgctggaggagctggtgctgaagctgcagcccc 1099
            | |||   ||    ||  ||| | |  ||  |    || ||   |||||||    |
Db   1124 cactgtttgccaatgcggaccccaacttcgtgacgtccatgctgaccaagctgcgtttcg 1183

Qy   1100 agacctactcaccaggtgaatatgtatgccgcaaaggagacattggccaagagatgtaca 1159
          ||  || |   || || || ||  |   |||  ||||   ||||||| |  ||||||||
Db   1184 aggtcttccagcctggggactacatcatccgggaaggcaccattggcaagaagatgtact 1243

Qy   1160 tcatccgagagggtcaactggccgtggtggcagatg 1195
          ||||||   | ||     |   |||| |  |  | |
Db   1244 tcatccagcatggcgtggtcagcgtgctcaccaagg 1279
```

RESULT 2
US-09-226-012-1
; Sequence 1, Application US/09226012
; Patent No. 6207383
; GENERAL INFORMATION:
; APPLICANT: Keating, Mark T.
; APPLICANT: Splawski, Igor
; TITLE OF INVENTION: MUTATIONS IN AND GENOMIC STRUCTURE OF HERG - A LONG QT
; TITLE OF INVENTION: SYNDROME GENE
; FILE REFERENCE: 2323-136
; CURRENT APPLICATION NUMBER: US/09/226,012
; CURRENT FILING DATE: 1999-01-06
; EARLIER APPLICATION NUMBER: 09/122,847
; EARLIER FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3480
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3477)
US-09-226-012-1

Query Match 3.5%; Score 59.8; DB 4; Length 3480;
Best Local Similarity 45.9%; Pred. No. 9.8e-07;
Matches 205; Conservative 0; Mismatches 242; Indels 0; Gaps 0;

```
Qy    726 agagtacctcttcatggtgggcgacttcctgctggccgtcatgggtttcgccaccatcat 785
          ||||  |  |||||    |  ||| | | ||  | | |  | |    |  || | |||| |
Db   1908 agagaagatcttctccatctgcgtcatgctcattggctccctcatgtatgctagcatctt 1967

Qy    786 gggtagcatgagctctgtcatctacaacatgaacactgcagatgcggctttctacccaga 845
          || | | ||    |  |||||  |     || || | |     ||    | | || || |
Db   1968 cggcaacgtgtcggccatcatccagcggctgtactcgggcacagcccgctaccacacaca 2027

Qy    846 tcatgcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggagcggcg 905
```

```
                | ||||  | |||  |||     | ||  ||  ||  |      |||   | |||
Db   2028 gatgctgcgggtgcgggagttcatccgcttccaccagatccccaatcccctgcgccagcg 2087

Qy    906 agttattgactggtatcagcacctgcagatcaacaagaagatgaccaacgaggtagccat 965
            |   || |  |  ||||||     |  | |||  ||    | | ||  |   ||  |
Db   2088 cctcgaggagtacttccagcacgcctggtcctacaccaacggcatcgacatgaacgcggt 2147

Qy    966 cttacagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctgtccac 1025
            |  ||  ||| ||||||  | |||| ||| ||  |     ||    || |||  ||
Db   2148 gctgaagggcttccctgagtgcctgcaggctgacatctgcctgcacctgaaccgctcact 2207

Qy   1026 tctgagccgggtgcagatctttcagaactgtgaggccagcctgctggaggagctggtgct 1085
           |||   |      |   ||| |                  ||   ||   ||  ||||   |
Db   2208 gctgcagcactgcaaacccttccgaggggccaccaagggctgccttcgggccctggccat 2267

Qy   1086 gaagctgcagccccagacctactcaccaggtgaatatgtatgccgcaaaggagacattgg 1145
          |||| |  || ||          | ||||| ||     |    |     || ||| |
Db   2268 gaagttcaagaccacacatgcaccgccaggggacacactggtgcatgctggggacctgct 2327

Qy   1146 ccaagagatgtacatcatccgagaggg 1172
          |   |  ||||| |||||     |||
Db   2328 caccgccctgtacttcatctcccgggg 2354
```

RESULT 3
US-09-226-012-3
; Sequence 3, Application US/09226012
; Patent No. 6207383
; GENERAL INFORMATION:
; APPLICANT: Keating, Mark T.
; APPLICANT: Splawski, Igor
; TITLE OF INVENTION: MUTATIONS IN AND GENOMIC STRUCTURE OF HERG - A LONG QT
; TITLE OF INVENTION: SYNDROME GENE
; FILE REFERENCE: 2323-136
; CURRENT APPLICATION NUMBER: US/09/226,012
; CURRENT FILING DATE: 1999-01-06
; EARLIER APPLICATION NUMBER: 09/122,847
; EARLIER FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 3950
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (67)..(3543)
US-09-226-012-3

Query Match 3.5%; Score 59.8; DB 4; Length 3950;
Best Local Similarity 45.9%; Pred. No. 1e-06;
Matches 205; Conservative 0; Mismatches 242; Indels 0; Gaps 0;

```
Qy    726 agagtacctcttcatggtgggcgacttcctgctggccgtcatgggtttcgccaccatcat 785
          ||||  |  |||||    |  ||| | | ||  | | |  | |    |  || | |||| |
Db   1974 agagaagatcttctccatctgcgtcatgctcattggctccctcatgtatgctagcatctt 2033

Qy    786 gggtagcatgagctctgtcatctacaacatgaacactgcagatgcggctttctacccaga 845
          || | | ||    |  |||||  |     || || | |     ||    | | || || |
Db   2034 cggcaacgtgtcggccatcatccagcggctgtactcgggcacagcccgctaccacacaca 2093

Qy    846 tcatgcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggagcggcg 905
                | ||||  | |||  |||     | ||  ||  ||  |      |||   | |||
Db   2094 gatgctgcgggtgcgggagttcatccgcttccaccagatccccaatcccctgcgccagcg 2153

Qy    906 agttattgactggtatcagcacctgcagatcaacaagaagatgaccaacgaggtagccat 965
            |   || |  |  ||||||     |  | |||  ||    | | ||  |   ||  |
Db   2154 cctcgaggagtacttccagcacgcctggtcctacaccaacggcatcgacatgaacgcggt 2213

Qy    966 cttacagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctgtccac 1025
          . |  ||  ||| ||||||  | |||| ||| ||  |     ||    || |||  ||
Db   2214 gctgaagggcttccctgagtgcctgcaggctgacatctgcctgcacctgaaccgctcact 2273
```

```
Qy   1026 tctgagccgggtgcagatctttcagaactgtgaggccagcctgctggaggagctggtgct 1085
          |||   |     |   ||| |              ||   ||   ||  ||||   |
Db   2274 gctgcagcactgcaaacccttccgaggggccaccaagggctgccttcgggccctggccat 2333

Qy   1086 gaagctgcagccccagacctactcaccaggtgaatatgtatgccgcaaaggagacattgg 1145
          |||| |  || ||          | ||||| ||     |    |     || ||| |
Db   2334 gaagttcaagaccacacatgcaccgccaggggacacactggtgcatgctggggacctgct 2393

Qy   1146 ccaagagatgtacatcatccgagaggg 1172
          |   |   ||||| |||||     |||
Db   2394 caccgccctgtacttcatctcccgggg 2420


RESULT   4
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-Fls
US-08-232-463-14


  Query Match             3.1%;  Score 52.8;  DB 1;  Length 7218;
  Best Local Similarity   2.4%;  Pred. No. 0.00011;
  Matches    9;  Conservative  223;  Mismatches  150;  Indels    0;  Gaps    0;

Qy   1340 aggaggacctgcgggaggtgctgagcgagtatccacaagcacagaccatcatggaggaga 1399
          | :::::   : :::::: :  ::: ::: :   : ::: : :::  :  : ::::::::
Db   1436 ACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1377

Qy   1400 aaggacgtgagatcctgctgaaaatgaacaagttggacgtgaatgctgaggcagctgaga 1459
          :::::: : ::::    :  ::::: ::: :::  ::: : ::: :  :::: ::  ::::
Db   1376 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1317

Qy   1460 tcgccctgcaggaggccacagagtcccggctacgaggcctagaccagcagctggatgatc 1519
            :    : ::::::  : ::::    ::  : ::::  :::  :: ::  ::: ::
Db   1316 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1257

Qy   1520 tacagaccaagtttgctcgcctcctggctgagctggagtccagcgcacttaagattgctt 1579
          : :::  :::   :  :       ::  :::  ::::   :: : :   ::::  :
Db   1256 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1197

Qy   1580 accgcattgaacggctggagtggcagactcgagagtggccaatgcccgaggacctggctg 1639
          :  : :  ::: ::  :::: :: :::   ::::: ::  :: :  :::::  ::  :
Db   1196 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1137

Qy   1640 aggctgatgacgagggtgagcctgaggagggaacttccaaagatgaagagggcagggcca 1699
          :::  :: :: ::::: :::   ::::::::::     ::::: :::::::: ::::  :
Db   1136 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1077

Qy   1700 gccaggagggacccccaggtcc 1721
          :  :::::::| | | || |||
Db   1076 RRRRRRRRRRATCGCAAGCTCC 1055


RESULT   5
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-Fls
US-08-232-463-14
```

```
  Query Match             3.0%;  Score 52.4;  DB 1;  Length 7218;
  Best Local Similarity   8.2%;  Pred. No. 0.00014;
  Matches   32;  Conservative  197;  Mismatches  163;  Indels    0;  Gaps    0;

Qy     314 tctcgagtcgctacgttcgcacctggagtttcttcttggacctggcttccctgatgccca 373
           ::::   :: :: : ::: : :::    :::::::::   :::  :::::::  : :::
Db    1105 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1164

Qy     374 cagatgtggtctacgtgcggctgggcccgcacacacccaccctgaggctgaaccgctttc 433
           :   : :  ::: : : :  ::   ::: : : : ::: ::::    ::   :: :::::
Db    1165 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1224

Qy     434 tccgcgcgccccgcctcttcgaggccttcgaccgcacagagacccgcacagcttacccaa 493
           ::: : : :::: ::::::::    :::::  :: : :      ::: : :  ::: :::
Db    1225 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1284

Qy     494 atgcctttcgcattgccaagctgatgctttacatttttgtcgtcatccattggaacagct 553
           : :::::: : :: ::   ::  : :::: : ::::: :: :: ::: ::    :  ::
Db    1285 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1344

Qy     554 gcctatactttgccctatcccggtacctgggcttcgggcgtgacgcatgggtgtacccgg 613
           ::: : :::: :::: ::::   : :::   ::::   : :  :  : : :  : : :::
Db    1345 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1404

Qy     614 accccgcgcagcctggctttgagcgcctgcggcgccagtacctctatagcttttacttct 673
           :::: : :  :::  ::::   : ::: :|   | | | | |||||  |||| ||| ||
Db    1405 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYGTACCAAATTCTTCTATCTCTTTAACTACT 1464

Qy     674 ccacgctgatactgactacagtgggcgataca 705
                     | | |||||||     ||||
Db    1465 TGCATAGATAGGTAATTACAGTGATGCCTACA 1496


RESULT   6
US-09-600-776-1
; Sequence 1, Application US/09600776
; Patent No. 6326168
; GENERAL INFORMATION:
;  APPLICANT:  Yamanouchi Pharmaceutical Co., Ltd.
;  TITLE OF INVENTION:  A novel potassium channel protein
;  FILE REFERENCE:  Y9903-PCT
;  CURRENT APPLICATION NUMBER: US/09/600,776
;  CURRENT FILING DATE:  2000-07-21
;  PRIOR APPLICATION NUMBER:  JP P1998-011434
;  PRIOR FILING DATE:  1998-01-23
;  PRIOR APPLICATION NUMBER:  JP P1998-346198
;  PRIOR FILING DATE:  1998-12-04
;  NUMBER OF SEQ ID NOS:  12
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 3323
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (6)..(3257)
US-09-600-776-1


  Query Match             3.0%;  Score 52;  DB 4;  Length 3323;
  Best Local Similarity   45.4%;  Pred. No. 0.00013;
  Matches  267;  Conservative    0;  Mismatches  315;  Indels    6;  Gaps    2;

Qy     611 cggaccccgcgcagcctggctttgagcgcctgcggcgccagtacctctatagcttttact 670
           |||  |  | || ||  |||     |  |||||  ||  |||  ||     | | ||||
Db    1318 cggggctggagctgctgggcggcccgtcgctgcgcagcgcctacatcacctccctctact 1377

Qy     671 tctccacgctgatactgactacagtggg---cgatacaccgccgccagccagggaagaag 727
           || |       |  || || |  |||||   ||  |    | |  |   || |||    |
Db    1378 tcgcactcagcagcctcaccagcgtgggcttcggcaacgtgtccgccaacacggacaccg 1437

Qy     728 agtacctcttcatggtgggcgacttcctgctggccgtcatgggtttcgccaccatcatgg 787
           || |  |||||    |  ||  | | ||  | |-||| | ||     |||     |  | |
Db    1438 agaagatcttctccatctgcaccatgctcatcggcgccctgatgcacgcggtggtgtttg 1497

Qy     788 gtagcatgagctctgtcatctacaacatgaacactgcagatgcggctttctacccagatc 847
           | | | |||    |  ||||| |   |||| || |             | | ||
Db    1498 ggaacgtgacggccatcatccagcgcatgtacgcccgccgctttctgtaccacagccgca 1557

Qy     848 atgcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggagcggcgag 907
                 ||    | |||||     | || |   ||  |      ||  ||| |||
Db    1558 cgcgcgacctgcgcgactacatccgcatccaccgtatccccaagcccctcaagcagcgca 1617

Qy     908 ttattgactggtatcagcacctgcagatcaacaagaagatgaccaacgaggtagccatct 967
           |  | || |  |  |||  |      |       || ||    | | ||       |  |
Db    1618 tgctggagtacttccaggccacctgggcggtgaacaatggcatcgacaccaccgagctgc 1677

Qy     968 tacagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctgtccactc 1027
           | |||  | | |||||   ||||| |||||  | ||  ||    |||||  |      |
Db    1678 tgcagagcctccctgacgagctgcgcgcagacatcgccatgcacctgcacaaggaggtcc 1737

Qy    1028 tgagccgggtgcagatctttcagaactgtgaggccagcctgctggaggagctggtgctga 1087
           ||   | | |||   | ||| ||           | ||   |||  ||  |||   |||
Db    1738 tg---cagctgccactgtttgaggcggccagccgcggctgcctgcgggcactgtctctgg 1794

Qy    1088 agctgcagccccagacctactcaccaggtgaatatgtatgccgcaaaggagacattggcc 1147
             |||| ||||     || | | || || || ||  |   || | |||| ||       |
Db    1795 ccctgcggcccgccttctgcacgccgggcgagtacctcatccaccaaggcgatgccctgc 1854

Qy    1148 aagagatgtacatcatccgagagggtcaactggccgtggtggcagatg 1195
           | |   | |||  |  || |    ||     |||  |||  |    | ||
Db    1855 aggccctctactttgtctgctctggctccatggaggtgctcaagggtg 1902


RESULT   7
US-09-600-776-5
; Sequence 5, Application US/09600776
; Patent No. 6326168
; GENERAL INFORMATION:
;  APPLICANT:  Yamanouchi Pharmaceutical Co., Ltd.
;  TITLE OF INVENTION:  A novel potassium channel protein
;  FILE REFERENCE:  Y9903-PCT
;  CURRENT APPLICATION NUMBER: US/09/600,776
;  CURRENT FILING DATE:  2000-07-21
;  PRIOR APPLICATION NUMBER:  JP P1998-011434
;  PRIOR FILING DATE:  1998-01-23
;  PRIOR APPLICATION NUMBER:  JP P1998-346198
;  PRIOR FILING DATE:  1998-12-04
;  NUMBER OF SEQ ID NOS:  12
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 3064
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (4)..(3057)
US-09-600-776-5


  Query Match             2.9%;  Score 49.8;  DB 4;  Length 3064;
  Best Local Similarity   45.5%;  Pred. No. 0.00051;
  Matches  217;  Conservative    0;  Mismatches  257;  Indels    3;  Gaps    1;

Qy     717 cagggaagaagagtacctcttcatggtgggcgacttcctgctggccgtcatgggtttcgc 776
           ||  || |  ||| |  |||||     |  ||    | ||  | | || | ||     |||
Db    1347 caccgacgcggagaagatcttctccatctgcacgatgctcataggcgccctgatgcacgc 1406

Qy     777 caccatcatgggtagcatgagctctgtcatctacaacatgaacactgcagatgcggcttt 836
             |  | || | | |||    |  |||||  |   |||| || |         ||   |
Db    1407 tgtggtgttcgggaacgtgacagccatcatccagcgcatgtactcgcgccgctcgctcta 1466

Qy     837 ctacccagatcatgcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaagct 896
           | ||                 | ||| | | |||     |||| | | |    ||   |||
Db    1467 ccacagccgcatgaaggacctcaaggacttcatccgtgtgcaccgcctgccgcggccgct 1526
```

```
Qy    897 ggagcggcgagttattgactggtatcagcacctgcagatcaacaagaagatgaccaacga 956
Db   1527 caagcagcgcatgctcgaatacttccagaccacgtgggccgtcaacagcggcatcgacgc 1586

Qy    957 ggtagccatcttacagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtgca 1016
Db   1587 caacgagttactgcgtgacttcccagacgagctgagagctgacattgctatgcacctgaa 1646

Qy   1017 cctgtccactctgagccgggtgcagatctttcagaactgtgaggccagcctgctggagga 1076
Db   1647 tcgggagatcctgcagctgccgttgttcggggcagcgagcaggggctgcctgcgggcc-- 1704

Qy   1077 gctggtgctgaagctgcagccccagacctactcaccaggtgaatatgtatgccgcaaagg 1136
Db   1705 -ctatcgctgcacatcaagacctcgttctgcgctccgggcgagtacctgttgcgccgtgg 1763

Qy   1137 agacattggccaagagatgtacatcatccgagagggtcaactggccgtggtggcaga 1193
Db   1764 ggatgccctgcaggcacattactatgtctgctccggctcgcttgaggtgctccgaga 1820
```

RESULT 8
US-08-956-242-1
; Sequence 1, Application US/08956242C
; Patent No. 5986081
; GENERAL INFORMATION:
; APPLICANT: Ganetzky, Barry S.
; APPLICANT: Titus, Steven A.
; TITLE OF INVENTION: Polynucleotides Encoding Herg-3
; FILE REFERENCE: 960296.94550
; CURRENT APPLICATION NUMBER: US/08/956,242C
; CURRENT FILING DATE: 1997-10-22
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3141
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (248)..(2128)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)
; OTHER INFORMATION: Unidentified at time of filing
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (3)
; OTHER INFORMATION: Unidentified at time of filing
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (12)
; OTHER INFORMATION: Unidentified at time of filing
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1568)..(1872)
; OTHER INFORMATION: Unidentified at time of filing
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (3126)
; OTHER INFORMATION: Unidentified at time of filing
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (3134)
; OTHER INFORMATION: Unidentified at time of filing
US-08-956-242-1

Query Match 2.8%; Score 49; DB 2; Length 3141;
Best Local Similarity 28.1%; Pred. No. 0.00085;
Matches 207; Conservative 0; Mismatches 520; Indels 9; Gaps 2;

```
Qy    641 tgcggcgccagtacctctatagcttttacttctccacgctgatactgactacagtggg-- 698
Db   1242 tgcaggacaagtatgtcacagccctctacttcaccttcagcagcctcaccagcgtgggct 1301

Qy    699 -cgatacaccgccgccagccagggaagaagagtacctcttcatggtgggcgacttcctgc 757
Db   1302 tcggcaatgtctcgcccaacaccaactccgagaaggtcttctccatctgcgtcatgctca 1361

Qy    758 tggccgtcatgggtttcgccaccatcatgggtagcatgagctctgtcatctacaacatga 817
Db   1362 tcggctccctgatgtacgccagcatcttcgggaacgtgtccgcgatcatccagcgcctgt 1421

Qy    818 acactgcagatgcggctttctacccagatcatgcactggtgaagaagtacatgaagctgc 877
Db   1422 actcgggaactgccaggtaccacatgcagatgctgcgagtaaaagagttcattcgctttc 1481

Qy    878 agcacgtcaaccgcaagctggagcggcgagttattgactggtatcagcacctgcagatca 937
Db   1482 accaaatccccaaccctctgaggcaacgtcttgaagaatatttccagcacgcatggactt 1541

Qy    938 acaagaagatgaccaacgaggtagccatcttacagcacttgcctgagcggctgcgggcag 997
Db   1542 acaccaatggcattgacatg------aacatgnnnnnnnnnnnnnnnnnnnnnnnnnnnn 1595

Qy    998 aagtggctgtgtctgtgcacctgtccactctgagccgggtgcagatctttcagaactgtg 1057
Db   1596 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 1655

Qy   1058 aggccagcctgctggaggagctggtgctgaagctgcagccccagacctactcaccaggtg 1117
Db   1656 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 1715

Qy   1118 aatatgtatgccgcaaaggagacattggccaagagatgtacatcatccgagagggtcaac 1177
Db   1716 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 1775

Qy   1178 tggccgtggtggcagatgatggtatcacacagtatgctgtgctcggtgcagggctctact 1237
Db   1776 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 1835

Qy   1238 ttggggagatcagcatcatcaacatcaaagggaacatgtctgggaaccgccgcacagcca 1297
Db   1836 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 1895

Qy   1298 acatcaagagcctaggttattcagacctattctgcctgagcaaggaggacctgcgggagg 1357
Db   1896 acgtgcgggctctgacctactgcgacctgcacaagatccagcgggcagatctgctggagg 1955

Qy   1358 tgctgagcgagtatcc 1373
Db   1956 tgctggacatgtaccc 1971
```

RESULT 9
US-09-351-215-1
; Sequence 1, Application US/09351215
; Patent No. 6087488
; GENERAL INFORMATION:
; APPLICANT: Ganetzky, Barry S.
; APPLICANT: Titus, Steven A.
; TITLE OF INVENTION: Polynucleotides Encoding Herg-3
; FILE REFERENCE: 960296.94550
; CURRENT APPLICATION NUMBER: US/09/351,215
; CURRENT FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: 08/956,242
; EARLIER FILING DATE: 1997-10-22
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3141
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:

```
;   NAME/KEY: CDS
;   LOCATION: (248)..(2128)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (3)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (12)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1568)..(1872)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (3126)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (3134)
;   OTHER INFORMATION: Unidentified at time of filing
US-09-351-215-1


  Query Match             2.8%;  Score 49;  DB 3;  Length 3141;
  Best Local Similarity   28.1%;  Pred. No. 0.00085;
  Matches  207;  Conservative    0;  Mismatches  520;  Indels    9;  Gaps    2;

Qy     641 tgcggcgccagtacctctatagcttttacttctccacgctgatactgactacagtggg-- 698
           ||| |  | ||||  ||     | | |||||| ||      |  || || |  |||||
Db    1242 tgcaggacaagtatgtcacagccctctacttcaccttcagcagcctcaccagcgtgggct 1301

Qy     699 -cgatacaccgccgccagccagggaagaagagtacctcttcatggtgggcgacttcctgc 757
            ||  |      ||||   ||   |    ||| |  |||||    |  ||| | | ||
Db    1302 tcggcaatgtctcgcccaacaccaactccgagaaggtcttctccatctgcgtcatgctca 1361

Qy     758 tggccgtcatgggtttcgccaccatcatgggtagcatgagctctgtcatctacaacatga 817
           | | |  | ||   | ||||| |||| | || | | ||  | |  ||||| |   | ||
Db    1362 tcggctccctgatgtacgccagcatcttcgggaacgtgtccgcgatcatccagcgcctgt 1421

Qy     818 acactgcagatgcggctttctacccagatcatgcactggtgaagaagtacatgaagctgc 877
           || | | |  |||    | | ||    |       |  || ||  ||| |||     | |
Db    1422 actcgggaactgccaggtaccacatgcagatgctgcgagtaaaagagttcattcgctttc 1481

Qy     878 agcacgtcaaccgcaagctggagcggcgagttattgactggtatcagcacctgcagatca 937
           | ||  ||  |  |   |||  ||  ||  ||   || |  |  ||||||     ||
Db    1482 accaaatccccaaccctctgaggcaacgtcttgaagaatatttccagcacgcatggactt 1541

Qy     938 acaagaagatgaccaacgaggtagccatcttacagcacttgcctgagcggctgcgggcag 997
           |||  ||    |   ||  |      | | |
Db    1542 acaccaatggcattgacatg------aacatgnnnnnnnnnnnnnnnnnnnnnnnnnnnn 1595

Qy     998 aagtggctgtgtctgtgcacctgtccactctgagccgggtgcagatctttcagaactgtg 1057

Db    1596 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 1655

Qy    1058 aggccagcctgctggaggagctggtgctgaagctgcagccccagacctactcaccaggtg 1117

Db    1656 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 1715

Qy    1118 aatatgtatgccgcaaaggagacattggccaagagatgtacatcatccgagagggtcaac 1177

Db    1716 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 1775

Qy    1178 tggccgtggtggcagatgatggtatcacacagtatgctgtgctcggtgcagggctctact 1237

Db    1776 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 1835

Qy    1238 ttggggagatcagcatcatcaacatcaaagggaacatgtctgggaaccgccgcacagcca 1297
                                                  |  ||  |     ||  ||
Db    1836 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnncacgaggggcaagtccagtgcag 1895

Qy    1298 acatcaagagcctaggttattcagacctattctgcctgagcaaggaggacctgcgggagg 1357
           || |   |   ||    || |  |||||   |    |      ||  || |||| |||||
Db    1896 acgtgcgggctctgacctactgcgacctgcacaagatccagcgggcagatctgctggagg 1955

Qy    1358 tgctgagcgagtatcc 1373
           |||||  |  ||| ||
Db    1956 tgctggacatgtaccc 1971


RESULT  10
US-08-979-608A-14
; Sequence 14, Application US/08979608A
; Patent No. 6355451
;  GENERAL INFORMATION:
;       APPLICANT: Lees, Ann M.
;                  Lees, Robert S.
;                  Law, Simon W.
;                  Arjona, Anibal A.
;       TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
;                           BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
;                           TREATING ATHEROSCLEROSIS
;       NUMBER OF SEQUENCES: 42
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Fish & Richardson P.C.
;            STREET: 225 Franklin Street
;            CITY: Boston
;            STATE: MA
;            COUNTRY: USA
;            ZIP: 02110-2804
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Diskette
;            COMPUTER: IBM Compatible
;            OPERATING SYSTEM: DOS
;            SOFTWARE: FastSEQ for Windows Version 2.0
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/08/979,608A
;            FILING DATE: 26-No. 6355451-1997
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US 60/048,547
;            FILING DATE: 03-JUN-1997
;            APPLICATION NUMBER: US 60/031,930
;            FILING DATE: 27-NOV-1996
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Myers, Louis
;            REGISTRATION NUMBER: 35,965
;            REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 617/542-5070
;            TELEFAX: 617/542-8906
;  INFORMATION FOR SEQ ID NO: 14:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 4722 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       FEATURE:
;            NAME/KEY:  Coding Sequence
;            LOCATION:  61...1731
;       SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-08-979-608A-14


  Query Match             2.5%;  Score 44;  DB 4;  Length 4722;
  Best Local Similarity   53.5%;  Pred. No. 0.024;
  Matches   92;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy    1381 cagaccatcatggaggagaaaggacgtgagatcctgctgaaaatgaacaagttggacgtg 1440
           | ||| | |   ||||    ||  || ||   || |  |  |  | | ||| | | |
Db     496 CGGACGAGCGATGAGGTCGGAGACCGAGACCACCGGAGGCCACAGGAAAAGAAGAAGGCC 555
```

```
Qy   1441 aatgctgaggcagctgagatcgccctgcaggaggccacagagtcccggctacgaggccta 1500
          || | |  || |  |||||| | |||| |  |   |||  |  |  |||  |   || |
Db    556 AAGGGTCTGGGAAAGGAGATCACGCTGCTGATGCAGACACTGAACACGCTGAGCACCCCA 615

Qy   1501 gaccagcagctggatgatctacagaccaagtttgctcgcctcctggctgagc 1552
          ||  || ||||||  | |||    |  |||| |||    || || |  ||||
Db    616 GAGGAGAAGCTGGCGGCTCTGTGCAAGAAGTATGCGGAACTGCTCGAGGAGC 667


RESULT  11
US-08-370-909-1
; Sequence 1, Application US/08370909
; Patent No. 5843648
;  GENERAL INFORMATION:
;    APPLICANT:  ROBBINS, PAUL F.; ROSENBERG,
;    APPLICANT:  STEVEN A.
;    TITLE OF INVENTION:  p15 AND TYROSINASE
;    TITLE OF INVENTION:  MELANOMA ANTIGENS AND THEIR USE IN DIAGNOSTIC
;    TITLE OF INVENTION:  AND THERAPEUTIC METHODS
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN, L.L.P.
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/370,909
;      FILING DATE:  10-JAN-1995
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CAROL M. GRUPPI
;      REGISTRATION NUMBER:  37,341
;      REFERENCE/DOCKET NUMBER:  2026-4155
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;      TELEX:  421792
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  809
;      TYPE:  NUCLEOTIDE
;      STRANDEDNESS:  DOUBLE
;      TOPOLOGY:  UNKNOWN
;    MOLECULE TYPE:  cDNA
US-08-370-909-1


  Query Match             2.4%;  Score 41.8;  DB 2;  Length 809;
  Best Local Similarity   46.0%;  Pred. No. 0.044;
  Matches  142;  Conservative    0;  Mismatches  167;  Indels    0;  Gaps    0;

Qy    802 gtcatctacaacatgaacactgcagatgcggctttctacccagatcatgcactggtgaag 861
          | | | |||| | | || ||  |  | | ||   | | | |  |   ||   ||| |  |
Db    184 GACTTTTACATCCTCAAGACCCCGAAGGAGGACCTGTGCTCCAAGTTTGGGATGGAGCTG 243

Qy    862 aagtacatgaagctgcagcacgtcaaccgcaagctggagcggcgagttattgactggtat 921
          |||     || ||||| ||   |   |||  ||      | ||        ||
Db    244 AAGCGAGGGATGCTGCTGCGGCTTGCCCGGCAGGACCCCCAGCTGCACCCCGAGGACCCC 303

Qy    922 cagcacctgcagatcaacaagaagatgaccaacgaggtagccatcttacagcacttgcct 981
            |||  | |     || | |  | | |  ||| ||  | ||||||  | || |   | |
Db    304 GAGCGGCGGGCAGCCATCTACGACAAGTACAAGGAATTTGCCATCCCAGAGGAGGAGGCA 363

Qy    982 gagcggctgcgggcagaagtggctgtgtctgtgcacctgtccactctgagccgggtgcag 1041
          ||| || || |        |||  | | |  |  |   |   |  ||      || | ||
Db    364 GAGTGGGTGGGCCTCACGCTGGAGGAGGCCATTGAGAAGCAGAGACTTTTGGAGGAGAAG 423

Qy   1042 atctttcagaactgtgaggccagcctgctggaggagctggtgctgaagctgcagccccag 1101
            |  |     |          | |    ||| ||||||| | | | |||||||||  |||
Db    424 GACCCTGTACCCCTGTTCAAGATCTATGTGGCGGAGCTGATCCAGCAGCTGCAGCAGCAG 483

Qy   1102 acctactca 1110
            |    |||
Db    484 GCACTGTCA 492


RESULT  12
US-09-056-105-22
; Sequence 22, Application US/09056105
; Patent No. 6287569
; GENERAL INFORMATION:
;  APPLICANT: KIPPS, THOMAS J.
;  APPLICANT: WU, YUNQI
;  TITLE OF INVENTION: VACCINES WITH ENHANCED INTRACELLULAR
;  TITLE OF INVENTION: PROCESSING
;  FILE REFERENCE: 233/221
;  CURRENT APPLICATION NUMBER: US/09/056,105
;  CURRENT FILING DATE: 1998-04-06
;  EARLIER APPLICATION NUMBER: 60/043,467
;  EARLIER FILING DATE: 1997-04-10
;  NUMBER OF SEQ ID NOS: 35
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 22
;   LENGTH: 809
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-056-105-22


  Query Match             2.4%;  Score 41.8;  DB 4;  Length 809;
  Best Local Similarity   46.0%;  Pred. No. 0.044;
  Matches  142;  Conservative    0;  Mismatches  167;  Indels    0;  Gaps    0;

Qy    802 gtcatctacaacatgaacactgcagatgcggctttctacccagatcatgcactggtgaag 861
          | | | |||| | | || ||  |  | | ||   | | | |  |   ||   ||| | |
Db    184 gacttttacatcctcaagaccccgaaggaggacctgtgctccaagtttgggatggagctg 243

Qy    862 aagtacatgaagctgcagcacgtcaaccgcaagctggagcggcgagttattgactggtat 921
          |||     || ||||| ||   |   |||  ||      | ||        ||
Db    244 aagcgagggatgctgctgcggcttgcccggcaggacccccagctgcaccccgaggacccc 303

Qy    922 cagcacctgcagatcaacaagaagatgaccaacgaggtagccatcttacagcacttgcct 981
            |||  | |     || | |  | | |  ||| ||  | ||||||  | || |   | |
Db    304 gagcggcgggcagccatctacgacaagtacaaggaatttgccatcccagaggaggaggca 363

Qy    982 gagcggctgcgggcagaagtggctgtgtctgtgcacctgtccactctgagccgggtgcag 1041
          ||| || || |        |||  | | |  |  |   |   |  ||      || | ||
Db    364 gagtgggtgggcctcacgctggaggaggccattgagaagcagagacttttggaggagaag 423

Qy   1042 atctttcagaactgtgaggccagcctgctggaggagctggtgctgaagctgcagccccag 1101
            |  |     |          | |    ||| ||||||| | | | |||||||||  |||
Db    424 gaccctgtacccctgttcaagatctatgtggcggagctgatccagcagctgcagcagcag 483

Qy   1102 acctactca 1110
            |    |||
Db    484 gcactgtca 492


RESULT  13
US-08-425-069-3
; Sequence 3, Application US/08425069
; Patent No. 5728810
;  GENERAL INFORMATION:
;    APPLICANT:  Lewis, Randolph V.
;    APPLICANT:  Xu, Ming
;    APPLICANT:  Hinman, Michael B.
;    TITLE OF INVENTION:  ISOLATED DNA CODING FOR SPIDER SILK
;    TITLE OF INVENTION:  PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
```

```
;    TITLE OF INVENTION:  CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
;    NUMBER OF SEQUENCES:  69
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;      STREET:  301 No. 5728810th Washington Street
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22046
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/425,069
;      FILING DATE:  19-APR-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy Jr., Gerald M
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1447-106P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 205-8000
;      TELEFAX:  (703) 205-8050
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1995 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    IMMEDIATE SOURCE:
;      CLONE:  p6B
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1785
US-08-425-069-3


  Query Match             2.4%;  Score 41;  DB 1;  Length 1995;
  Best Local Similarity   47.8%;  Pred. No. 0.11;
  Matches  119;  Conservative    0;  Mismatches  130;  Indels    0;  Gaps    0;

Qy     853 ctggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggagcggcgagttatt 912
           | ||  || ||| ||     |   |||   ||  | |   ||| |  || ||  |   |
Db     704 CAGGACAACAAGGACTATCTGGACCAGGCAGTGCAGCTGCAGCAGCCGCAGCAGGACCTG 763

Qy     913 gactggtatcagcacctgcagatcaacaagaagatgaccaacgaggtagccatcttacag 972
           |||     |  | |    | | ||  || || | |  |   || |  | | |    ||  |
Db     764 GACAACAAGGACCCGGAGGATATGGACCAGGACAACAAGGACCATCTGGACCCGGTAGTG 823

Qy     973 cacttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctgtccactctgagc 1032
           |   |||   ||| || || |    | |     | |  | ||  |||| ||    |  |
Db     824 CCGCTGCAGCAGCAGCCGCCGCAGCAGGACCTGGAGGATATGGCCCTGGACAACAAGGAC 883

Qy    1033 cgggtgcagatctttcagaactgtgaggccagcctgctggaggagctggtgctgaagctg 1092
           | || | | ||    ||| ||    ||| |   |||  | |||   ||  || | ||| |
Db     884 CCGGAGGATATGGACCAGGACAACAAGGACCATCTGGAGCAGGCAGTGCAGCAGCAGCAG 943

Qy    1093 cagccccag 1101
           | ||  |||
Db     944 CCGCAGCAG 952


RESULT  14
US-08-317-844B-3
; Sequence 3, Application US/08317844B
; Patent No. 5989894
;  GENERAL INFORMATION:
;    APPLICANT:  Lewis, Randolph V.
;    APPLICANT:  Xu, Ming
;    APPLICANT:  Hinman, Michael B.
;    TITLE OF INVENTION:  ISOLATED DNA CODING FOR SPIDER SILK
;    TITLE OF INVENTION:  PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
;    TITLE OF INVENTION:  CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
;    NUMBER OF SEQUENCES:  62
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;      STREET:  301 No. 5989894th Washington Street
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22046
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/317,844B
;      FILING DATE:  04-OCT-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy Jr., Gerald M
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1447-105P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 241-1300
;      TELEFAX:  (703) 241-2848
;      TELEX:  248345
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1995 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    IMMEDIATE SOURCE:
;      CLONE:  p6B
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1785
US-08-317-844B-3


  Query Match             2.4%;  Score 41;  DB 2;  Length 1995;
  Best Local Similarity   47.8%;  Pred. No. 0.11;
  Matches  119;  Conservative    0;  Mismatches  130;  Indels    0;  Gaps    0;

Qy     853 ctggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggagcggcgagttatt 912
           | ||  || ||| ||     |   |||   ||  | |   ||| |  || ||  |   |
Db     704 CAGGACAACAAGGACTATCTGGACCAGGCAGTGCAGCTGCAGCAGCCGCAGCAGGACCTG 763

Qy     913 gactggtatcagcacctgcagatcaacaagaagatgaccaacgaggtagccatcttacag 972
           |||     |  | |    | | ||  || || | |  |   || |  | | |    ||  |
Db     764 GACAACAAGGACCCGGAGGATATGGACCAGGACAACAAGGACCATCTGGACCCGGTAGTG 823

Qy     973 cacttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctgtccactctgagc 1032
           |   |||   ||| || || |    | |     | |  | ||  |||| ||    |  |
Db     824 CCGCTGCAGCAGCAGCCGCCGCAGCAGGACCTGGAGGATATGGCCCTGGACAACAAGGAC 883

Qy    1033 cgggtgcagatctttcagaactgtgaggccagcctgctggaggagctggtgctgaagctg 1092
           | || | | ||    ||| ||    ||| |   |||  | |||   ||  || | ||| |
Db     884 CCGGAGGATATGGACCAGGACAACAAGGACCATCTGGAGCAGGCAGTGCAGCAGCAGCAG 943

Qy    1093 cagccccag 1101
           | ||  |||
Db     944 CCGCAGCAG 952


RESULT  15
US-09-050-863-2
```

```
; Sequence 2, Application US/09050863
; Patent No. 6114111
;  GENERAL INFORMATION:
;    APPLICANT:  Lao, Ying
;    APPLICANT:  Hiang, Betty
;    APPLICANT:  Payan, Don
;    TITLE OF INVENTION:  Mammalian Protein Interaction Cloning
;    TITLE OF INVENTION:  System
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/050,863
;      FILING DATE:  30-MAR-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silva, Robin M.
;      REGISTRATION NUMBER:  38,304
;      REFERENCE/DOCKET NUMBER:  A-65638/DJB/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 949-8711
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2580 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA
US-09-050-863-2


  Query Match             2.3%;  Score 40.4;  DB 3;  Length 2580;
  Best Local Similarity   45.9%;  Pred. No. 0.18;
  Matches  177;  Conservative    0;  Mismatches  206;  Indels    3;  Gaps    1;

Qy   1340 aggaggacctgcgggaggtgctgagcgagtatccacaagcacagaccatcatggaggaga 1399
          || ||||   | || ||| || |   |||   |   | |  |||        ||||| |
Db    812 AGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGC 871

Qy   1400 aaggacgtgagatcctgctgaaaatgaacaagttggacgtgaatgctgaggcagctgaga 1459
          | |   |  ||     || | |   | |  ||  | | | | | |  ||||     |||
Db    872 AGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGG 931

Qy   1460 tcgccctgcaggaggccacagagtcccggctacgaggcctagaccagcagctggatgatc 1519
                |||||||       ||    |||      || |  || ||| ||  ||   |
Db    932 GGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGG 991

Qy   1520 tacagaccaagtttgctcgcctcctggctgagctggagtccagcgcacttaagattgctt 1579
            |||     |   |   |    | ||  | || ||||     |  |   ||   |
Db    992 AGCAGGAGGGGCAGGAGGG---GCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGG 1048

Qy   1580 accgcattgaacggctggagtggcagactcgagagtggccaatgcccgaggacctggctg 1639
          |  |    ||  ||| |||| |||||   |  ||| |||    ||  ||||  | ||
Db   1049 AGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGC 1108

Qy   1640 aggctgatgacgagggtgagcctgaggagggaacttccaaagatgaagagggcagggcca 1699
          |||  |   | |||   |||    |||||||        | || |   |||   ||
Db   1109 AGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGG 1168

Qy   1700 gccaggagggacccccaggtccagag 1725
            |||||||| |     || |  |||
Db   1169 AGCAGGAGGGGCAGGAGGGGCAGGAG 1194
```

Search completed: June 20, 2002, 14:28:41
Job time: 18644 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2002, 12:16:50 ; Search time 4682.41 Seconds
(without alignments)
4980.926 Million cell updates/sec

Title: US-09-927-267-3
Perfect score: 1728
Sequence: 1 atgagccaggacaccaaagt.....gacccccaggtccagagtga 1728

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	512.6	29.7	793	10	BI830763	BI830763 603075162
	2	328.6	19.0	646	9	BB625423	BB625423 BB625423
	3	246	14.2	699	12	CNS04SGS	AL305173 Tetraodon
c	4	243.6	14.1	800	12	CNS03HKC	AL244389 Tetraodon
	5	241.4	14.0	627	12	AQ537904	AQ537904 RPCI-11-3
	6	230.8	13.4	559	10	BF190904	BF190904 237569 MA
	7	227.4	13.2	560	10	BF190899	BF190899 237563 MA
	8	221.6	12.8	558	9	AW640105	AW640105 bl91d05.w
c	9	217.8	12.6	621	12	AZ503692	AZ503692 1M0343L05
	10	215.6	12.5	567	10	BF937274	BF937274 fm62h03.y
	11	204.4	11.8	819	12	CNS02XOL	AL218622 Tetraodon
	12	192.4	11.1	882	10	BG293093	BG293093 602389921
c	13	179.6	10.4	923	12	CNS042AH	AL271250 Tetraodon
c	14	167.4	9.7	580	10	BE491188	BE491188 db43c03.x
	15	150.8	8.7	502	10	W84514	W84514 zd90g10.r1
	16	134.6	7.8	378	12	AQ316849	AQ316849 RPCI11-97
	17	132.8	7.7	449	10	W58136	W58136 zd18b02.r1
	18	127.8	7.4	499	9	AW638930	AW638930 bl76e03.w
	19	126.6	7.3	803	10	BG473313	BG473313 602515428
	20	125.6	7.3	1019	10	BG342539	BG342539 602374710
c	21	115.6	6.7	568	10	W26690	W26690 11e4 Human
	22	105.8	6.1	454	9	AI035456	AI035456 ub47c01.r
	23	101.2	5.9	563	10	BG304577	BG304577 fl88b05.y
	24	100.6	5.8	515	10	BI706445	BI706445 fq06b04.y
c	25	100.4	5.8	743	10	BJ151852	BJ151852 BJ151852
	26	95.8	5.5	1048	12	CNS025BE	AL181859 Tetraodon
	27	91.2	5.3	747	10	BI733974	BI733974 603355503
	28	90	5.2	726	12	CNS01GHZ	AL143160 Anopheles
	29	89.2	5.2	573	10	BM354832	BM354832 rr17e02.y
c	30	85.2	4.9	518	9	AV605886	AV605886 AV605886
	31	83.2	4.8	376	10	T97907	T97907 ye54c11.r1
	32	83.2	4.8	698	10	BI548103	BI548103 603189424
c	33	76.2	4.4	471	10	BI506458	BI506458 BB170020B
c	34	76.2	4.4	471	10	BI509079	BI509079 BB170003A
c	35	75.8	4.4	410	12	BH121161	BH121161 RPCI-24-3
c	36	74.2	4.3	747	12	CNS04E7T	AL286706 Tetraodon
	37	70.6	4.1	694	9	AU137778	AU137778 AU137778
c	38	66.4	3.8	613	10	BI506874	BI506874 BB170026A
	39	62.4	3.6	353	9	AL120448	AL120448 DKFZp761I
	40	62.4	3.6	376	9	AL120239	AL120239 DKFZp761H
	41	60.4	3.5	881	10	BE287002	BE287002 601097229
	42	58.6	3.4	533	10	BJ123339	BJ123339 BJ123339
c	43	58.6	3.4	557	10	BI134074	BI134074 UI-M-BH3-
	44	56.6	3.3	926	10	BF315250	BF315250 601899133
	45	56.2	3.3	815	10	BE546989	BE546989 601071965

ALIGNMENTS

RESULT 1
BI830763
LOCUS BI830763 793 bp mRNA linear EST 04-OCT-2001
DEFINITION 603075162F1 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5167238 5', mRNA sequence.
ACCESSION BI830763
VERSION BI830763.1 GI:15942313
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 793)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11415 row: k column: 15
High quality sequence stop: 723.
FEATURES Location/Qualifiers
source 1. .793
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5167238"
/clone_lib="NIH_MGC_119"
/tissue_type="medulla"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source normal medulla from anonymous male age 27. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.3 kb, insert size range

0.9-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 013. Note: this is a NIH_MGC Library."
BASE COUNT 160 a 222 c 232 g 179 t
ORIGIN

Query Match 29.7%; Score 512.6; DB 10; Length 793;
Best Local Similarity 93.4%; Pred. No. 4.2e-114;
Matches 634; Conservative 0; Mismatches 34; Indels 11; Gaps 9;

```
Qy      60 caggaagttgctgcctgtcctggacccatctggggattactactactggtggctgaacac 119
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     115 CAGGAAGTTGCTGCCTGTCCTGGACCCATCTGGGGATTACTACTACTGGTGGCTGAACAC 174

Qy     120 aatggtcttcccagtcatgtataacctcatcatcctcgtgtgcagagcctgcttccccga 179
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     175 AATGGTCTTCCCAGTCATGTATAACCTCATCATCCTCGTGTGCAGAGCCTGCTTCCCCGA 234

Qy     180 cttgcagcacggttatctggtggcctggttggtgctggactacacgagtgacctgctata 239
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     235 CTTGCAGCACGGTTATCTGGTGGCCTGGTTGGTGCTGGACTACACGAGTGACCTGCTATA 294

Qy     240 cctactagacatggtggtgcgcttccacacaggattcttggaacagggcatcctggtggt 299
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     295 CCTACTAGACATGGTGGTGCGCTTCCACACAGGATTCTTGGAACAGGGCATCCTGGTGGT 354

Qy     300 ggacaagggtaggatctcgagtcgctacgttcgcacctggagtttcttc-ttggacctgg 358
           |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db     355 GGACAAGGGTAGGATCTCGAGTCGCTACGTTCGCACCTGGAGTTTCTTCTTTGGACCTGG 414

Qy     359 cttccctgatgcccacagatgtggtctacg-tgcggctgggcccgcacac-acccaccct 416
           |||||||||||||||||||||||||||||| |||||||||||||| |||| || ||||||
Db     415 CTTCCCTGATGCCCACAGATGTGGTCTACGTTGCGGCTGGGCCCGTACACAACACACCCT 474

Qy     417 gaggctgaaccgctttctccgcgcgccccgcctcttcgaggccttcgaccgcacagagac 476
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     475 GAGGCTGAACCGCTTTCTCCGCGCGCCCCGCCTCTTCGAGGCCTTCGACCGCACAGAGAC 534

Qy     477 ccgcacagcttacccaaatgcctttcgcattgccaagctgatgctttacatttttgtcgt 536
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db     535 CCGCACAGCTTACCCAAATGCCTTTCGCATTGCCAAGCTGATGCTTTACATTATGTCGTC 594

Qy     537 catccattggaacagctgccta-tactttgccctatcccggtacctgggcttcgggcg-t 594
           |  | | | |||||||||| |||||||||||||||||||||||||||||||||||||| |
Db     595 ATCCATTGGAAGCAGCTGCCTATTACTTTGCCCTATCCCGGTACCTGGGCTTCGGGCGTT 654

Qy     595 gacgcatgggtgtacccggaccccg--cgcagcctggctttgagcgcctgc-ggcgccag 651
           |||||||||||||||||| ||||       | |||||||||||| |||| ||||||||
Db     655 GACGCATGGGTGTACCCGGAACCCGACGCAGCCTTGGCTTTGAGCGACTGCGGGCGCCAG 714

Qy     652 tacctctatagcttttacttctccacgctg-atactgactacagtgggcg--atacaccg 708
           |||||||||||||||||||||||||||||| ||||||||||||||||||      ||||
Db     715 TACCTCTATAGCTTTTACTTCTCCACGCTGAATACTGACTACAGTGGGCGGATACAACCG 774

Qy     709 ccgccagccagggaagaag 727
           || ||  |||| ||||||
Db     775 ACGGCAAGCAGGAAAGAAG 793
```

RESULT 2
BB625423
LOCUS BB625423 646 bp mRNA linear EST 26-OCT-2001
DEFINITION BB625423 RIKEN full-length enriched, adult male epididymis Mus musculus cDNA clone 9230103C20 5', mRNA sequence.
ACCESSION BB625423
VERSION BB625423.1 GI:16463693
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 646)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.

FEATURES Location/Qualifiers
source 1. .646
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="9230103C20"
/clone_lib="RIKEN full-length enriched, adult male epididymis"
/sex="male"
/tissue_type="epididymis"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:

```
                    BamHI"
BASE COUNT      146 a    221 c    170 g    109 t
ORIGIN


  Query Match             19.0%;  Score 328.6;  DB 9;  Length 646;
  Best Local Similarity   79.6%;  Pred. No. 2.4e-69;
  Matches  413;  Conservative    0;  Mismatches  104;  Indels    2;  Gaps    2;

Qy      1 atgagccaggacaccaaagtgaagacaacagagtccagtcccccagccccatccaaggcc 60
          ||||||||||||| ||||||||||||||||||||||  ||||||||||| |||| |||
Db    127 ATGAGCCAGGACAGCAAAGTGAAGACAACAGAGTCCACGCCCCCAGCCCCAACCAAAGCC 186

Qy     61 aggaagttgctgcctgtcctggacccatctggggattactactactggtggctgaacaca 120
          ||||||  |||||| ||||| ||||||||||||||||| |||||||| ||||||||||
Db    187 AGGAAGCGGCTGCCAGTCCTAGACCCATCTGGGGATTATTACTACTGGGGGCTGAACACG 246

Qy    121 atggtcttcccagtcatgtataacctcatcatcctcgtgtgcagagcctgcttccccgac 180
          | || ||||||| |||||||||| |||||||||   |  ||||| |||||||| || |||
Db    247 AGGGCCTTCCCAATCATGTATAATCTCATCATCGCTGCATGCAGGGCCTGCTTTCCTGAC 306

Qy    181 ttgcagcacggttatctggtggcctggttggtgctggactacacgagtgacctgctatac 240
          ||||| | | | || |||||||||||| | | |||||||||||||||| |||||||| |||
Db    307 TTGCAACCCAGATACCTGGTGGCCTGGCTTGGGCTGGACTACACGAGGGACCTGCTGTAC 366

Qy    241 ctactagacatggtggtgcgcttccacacaggattcttggaacagggca-tcctggtggt 299
          ||||||||  | | ||| ||||||||||||||||||| | |||||||||| |||||||||
Db    367 CTACTAGATGTTGGGGTACGCTTCCACACAGGATTCCTAGAACAGGGCACTCCTGGTGGG 426

Qy    300 ggacaagggtaggatctcgagtcgctacgttcgcacctggagtttcttcttggacctggc 359
          ||||||  | | |||| |  | ||||| | |||| || ||||  || |  |||||| |||
Db    427 GGACAAAAGCACGATCGCCCGGCGCTATGGTCGCCCCCGGAGCCTCCTGGTGGACCCGGC 486

Qy    360 ttccctgatgcccacagatgtggtctacgtgcggctgggcccgcacacacccaccctgag 419
          ||||||| | ||| ||||||  | |||||||| | | |||||||||| |  ||| |  |
Db    487 TTCCCTGGTCCCCCCAGATGCAGCCTACGTGCAGTTAGGCCCGCACACCCGTACCGTCCG 546

Qy    420 gctgaaccgctttctccgcgcgccc-cgcctcttcgaggccttcgaccgcacagagaccc 478
          ||| ||||  || | ||||  |||| |||||||  | ||    || ||| || | |||||
Db    547 GCTAAACCCGTTCCCCCGCAGGCCCTCGCCTCTGTGCGGGTCCCGCCCGGACCGGGACCC 606

Qy    479 gcacagcttacccaaatgcctttcgcattgccaagctga 517
          | || ||| ||||| ||||| | |||  | ||||  |||
Db    607 GAACCGCTCACCCACATGCCCTCCGCTATACCAACGTGA 645


RESULT  3
CNS04SGS
LOCUS       CNS04SGS                699 bp    DNA     linear   GSS 24-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            007004 of library H from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL305173
VERSION     AL305173.1  GI:8196132
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 699)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 699)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 699)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .699
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="007004"
                     /clone_lib="H"
                     /note="Genoscope sequence ID : C0BH007BH02XE1-end :
                     PUC-Ori"
BASE COUNT      162 a    216 c    151 g    166 t      4 others
ORIGIN


  Query Match             14.2%;  Score 246;  DB 12;  Length 699;
  Best Local Similarity   59.4%;  Pred. No. 3.2e-49;
  Matches  411;  Conservative    3;  Mismatches  278;  Indels    0;  Gaps    0;

Qy    270 aggattcttggaacagggcatcctggtggtggacaagggtaggatctcgagtcgctacgt 329
          ||| ||  |:||||| || || ||||||  |||| | |  | | | | |    ||||: |
Db      8 AGGCTTTCTRGAACAAGGAATGCTGGTGAAGGACCAAGCAAAGCTGTGGGCCAGCTAMAT 67

Qy    330 tcgcacctggagtttcttcttggacctggcttccctgatgcccacagatgtggtctacgt 389
          || |||      |||    ||||| ||   ||| |   |||||| ||  ||| |||| |
Db     68 CCGAACCCACCAGTTCAAACTGGACTTGCTGTCCATCCAGCCCACTGAMCTGGCCTACAT 127

Qy    390 gcggctgggcccgcacacacccaccctgaggctgaaccgctttctccgcgcgccccgcct 449
             ||||  ||||  ||    || ||| | |||||| | ||||||   || ||| |
Db    128 CTCCACGGGCATCCACAAGCCGCAGCTCAGGTTCAACCGCCTCCTCCGCTTCCCGCGCAT 187

Qy    450 cttcgaggccttcgaccgcacagagacccgcacagcttacccaaatgcctttcgcattgc 509
          |||||  |||||| ||||| ||||| |||||    ||||| ||  ||| || |||
Db    188 GTTCGAATTCTTCGATCGCACCGAGACTCGCACCAACTACCCCAACATCTTCCGTATTTG 247

Qy    510 caagctgatgctttacatttttgtcgtcatccattggaacagctgcctatactttgccct 569
          ||| ||| |||| ||||| | ||| |||| || |||||  |||| |||| |  ||| |
Db    248 CAACCTGGTGCTCTACATCCTGGTCATCATTCACTGGAACGCCTGCATATATTACGCCAT 307

Qy    570 atcccggtacctgggcttcgggcgtgacgcatgggtgtacccggaccccgcgcagcctgg 629
          ||||  ||   |||| |||||    ||  | ||||| | |||  |     |    || |
Db    308 ATCCAAGTCTTTGGGGTTCGGTTCGGATACTTGGGTTTTCCCAAATATTTCCATACCGGA 367

Qy    630 ctttgagcgcctgcggcgccagtacctctatagcttttacttctccacgctgatactgac 689
          | |     ||||   ||    ||: |||| || | |||| || || || |  || ||
Db    368 ATATTCCTCCCTGACTCGGAGCTAMGTCTACTGCCTCTACTGGTCGACTCTTACTCTCAC 427

Qy    690 tacagtgggcgatacaccgccgccagccagggaagaagagtacctcttcatggtgggcga 749
          ||  | || || |  ||  | || | | | || ||||||||||||||  ||||    ||
Db    428 CACCATTGGAGAGATGCCCGCCCCCGTCCGAGACGAAGAGTACCTCTTTGTGGTCTTTGA 487

Qy    750 cttcctgctggccgtcatgggtttcgccaccatcatgggtagcatgagctctgtcatcta 809
          || ||  | | |||  ||   |||||||| ||  |||| | | |  |||  | ||
Db    488 TTTTCTCGTCGGCGTGCTGATCTTCGCCACAATTGTGGGGAACGTCGGCTCCATGATTGC 547

Qy    810 caacatgaacactgcagatgcggctttctacccagatcatgcactggtgaagaagtacat 869
          |||||||||| |  |    || |  ||| |  |       |      | ||  | |||||
Db    548 CAACATGAACGCCACCCGAGCCGAGTTCCAGGCTCGGATCGACGCCATCAAACACTACAT 607

Qy    870 gaagctgcagcacgtcaaccgcaagctggagcggcgagttattgactggtatcagcacct 929
          | |  | |   | ||||||    |||||||    || || ||  | ||||   |  ||||
Db    608 GCACTTCCGAAAAGTCAACAAAGAGCTGGAAACTCGCGTCATCAAGTGGTTCGACTACCT 667

Qy    930 gcagatcaacaagaagatgaccaacgaggtag 961
            || |||||||||        |||||  ||
Db    668 CTGGACCAACAAGAAAGCTGTGGACGAGCAAG 699
```

```
RESULT   4
CNS03HKC/c
LOCUS       CNS03HKC                800 bp    DNA     linear   GSS 17-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            027I05 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL244389
VERSION     AL244389.1  GI:7965401
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 800)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 800)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 800)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .800
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="027I05"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG027AE03SP1-end :
                     PUC-Ori"
BASE COUNT      173 a    228 c    195 g    199 t      5 others
ORIGIN
```

```
  Query Match             14.1%;  Score 243.6;  DB 12;  Length 800;
  Best Local Similarity   62.1%;  Pred. No. 1.3e-48;
  Matches  384;  Conservative    0;  Mismatches  234;  Indels    0;  Gaps    0;

Qy     832 gctttctacccagatcatgcactggtgaagaagtacatgaagctgcagcacgtcaaccgc 891
           ||||| |   || ||  || |    |       |        ||   | | | |||| |
Db     711 GCTTTTTTTTCATATTTGAATCTTTTTCTTCGGGTCCACAGGCCCGGAAGGTCAGCAAG 652

Qy     892 aagctggagcggcgagttattgactggtatcagcacctgcagatcaacaagaagatgacc 951
             | || |||  |||||| ||  | |||| | |  |  ||  || ||||   ||
Db     651 GACCTCGAGTTGCGAGTCATCAAGTGGTTTGACTATTTGTGGAACAACGGAAAAGCCCAG 592

Qy     952 aacgaggtagccatcttacagcacttgcctgagcggctgcgggcagaagtggctgtgtct 1011
             | |||  ||   |  |   | || | || ||   |||   ||| ||  | ||  |
Db     591 GATGAGAGAGAGGTGCTGAGGTACCTTCCGGACAAGCTCAAGGCGGAGATCGCCATCCAG 532

Qy    1012 gtgcacctgtccactctgagccgggtgcagatctttcagaactgtgaggccagcctgctg 1071
           || ||| ||   |||||||     || |  || |||    |||| |||||  ||||| ||
Db     531 GTCCACATGGATACTCTGAAGAAAGTCCGTATTTTTTCCGACTGCGAGGCGGGCCTGTTG 472

Qy    1072 gaggagctggtgctgaagctgcagccccagacctactcaccaggtgaatatgtatgccgc 1131
             |||||||||||| |||||||  |||  |||   | |   || || || ||  | |||
```

```
Db     471 ATCGAGCTGGTGCTCAAGCTGCGGCCGCAGGTGTTCAGCCCCGGGGACTACATCTGCAAG 412

Qy    1132 aaaggagacattggccaagagatgtacatcatccgagagggtcaactggccgtggtggca 1191
           || || ||||||||||  ||||||||||||||   || ||  | ||||| || || ||
Db     411 AAGGGCGACATTGGCCGTGAGATGTACATCATCAAGGACGGAAAGCTGGCGGTTGTCGCC 352

Qy    1192 gatgatggtatcacacagtatgctgtgctcggtgcagggctctactttggggagatcagc 1251
           |||||||| |||| || | ||  ||||| ||    ||    || || || |||||||||
Db     351 GATGATGGTGTCACGCAATTTGTGGTGCTGGGAAGTGGCAGTTATTTCGGCGAGATCAGC 292

Qy    1252 atcatcaacatcaaagggaacatgtctgggaaccgccgcacagccaacatcaagagccta 1311
           ||| | |||||||||||| | || | | |||||||| || || || ||||||   ||| |
Db     291 ATCCTTAACATCAAAGGCAGCAAGGCCGGGAACCGTCGAACGGCGAACATCCGCAGCATC 232

Qy    1312 ggttattcagacctattctgcctgagcaaggaggacctgcgggaggtgctgagcgagtat 1371
           ||||| || ||||| |||||||||   |||||| ||||||  ||||    ||   ||||||
Db     231 GGTTACTCGGACCTCTTCTGCCTCTCCAAGGACGACCTGATGGAGTCTTTGTTGGAGTAT 172

Qy    1372 ccacaagcacagaccatcatggaggagaaaggacgtgagatcctgctgaaaatgaacaag 1431
           ||  | ||  |  ||||  |||||||||| || ||  | |||||| ||||         |
Db     171 CCGGATGCCAAAGCCATGCTGGAGGAGAAGGGCCGGCAAATCCTGATGAAGGACGGTCTG 112

Qy    1432 ttggacgtgaatgctgag 1449
           |||||| || |  | | |
Db     111 TTGGACCTGGACCCGGCG 94
```

```
RESULT   5
AQ537904
LOCUS       AQ537904                627 bp    DNA     linear   GSS 18-MAY-1999
DEFINITION  RPCI-11-326D9.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-326D9,
            DNA sequence.
ACCESSION   AQ537904
VERSION     AQ537904.1  GI:4849594
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 627)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI-11-326D9.TV
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .627
                     /organism="Homo sapiens"
                     /db_xref="GDB:7624880"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-326D9"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      151 a    164 c    157 g    155 t
```

ORIGIN

Query Match 14.0%; Score 241.4; DB 12; Length 627;
Best Local Similarity 61.9%; Pred. No. 4e-48;
Matches 383; Conservative 0; Mismatches 236; Indels 0; Gaps 0;

```
Qy    454 gaggccttcgaccgcacagagacccgcacagcttacccaaatgcctttcgcattgccaag 513
          ||   ||| ||||||||||||||  | ||    ||||| |||   ||  | ||||  ||
Db      8 GAATTCTTTGACCGCACAGAGACAAGGACCAACTACCCCAATATGTTCAGGATTGGTAAC 67

Qy    514 ctgatgctttacatttttgtcgtcatccattggaacagctgcctatactttgccctatcc 573
          || |  | ||||||  |  || |||||||  |||||   |||| | |||||||||  | |||
Db     68 TTGGTCTTGTACATTCTCATCATCATCCACTGGAATGCCTGCATCTACTTTGCCATTTCC 127

Qy    574 cggtacctgggcttcgggcgtgacgcatgggtgtacccggaccccgcgcagcctggcttt 633
          || | | || || |||   ||| | ||||| |||||  ||  | |    || |    |
Db    128 AAGTTCATTGGTTTTGGGACAGACTCCTGGGTCTACCCAAACATCTCAATCCCAGAGCAT 187

Qy    634 gagcgcctgcggcgccagtacctctatagcttttacttctccacgctgatactgactaca 693
          | ||||||     |  ||||| | || ||  | ||||  |||||  |||  || || ||
Db    188 GGGCGCCTCTCCAGGAAGTACATTTACAGTCTCTACTGGTCCACCTTGACCCTTACCACC 247

Qy    694 gtgggcgatacaccgccgccagccagggaagaagagtacctcttcatggtgggcgacttc 753
          | || || || || || || |  |  || || ||||| |||||  |||| |  ||||||
Db    248 ATTGGTGAGACCCCACCCCCCGTGAAAGATGAGGAGTATCTCTTTGTGGTCGTAGACTTC 307

Qy    754 ctgctggccgtcatgggtttcgccaccatcatgggtagcatgagctctgtcatctacaac 813
          || |||  ||  ||  ||| |||||||  |||| |   || ||||  | ||||  ||
Db    308 TTGGTGGGTGTTCTGATTTTTGCCACCATTGTGGGCAATGTGGGCTCCATGATCTCGAAT 367

Qy    814 atgaacactgcagatgcggctttctacccagatcatgcactggtgaagaagtacatgaag 873
          |||||  |  ||   || |  ||| |  |  |   ||      | ||| |||||||| ||
Db    368 ATGAATGCCTCACGGGCAGAGTTCCAGGCCAAGATTGATTCCATCAAGCAGTACATGCAG 427

Qy    874 ctgcagcacgtcaaccgcaagctggagcggcgagttattgactggtatcagcacctgcag 933
          | |   | |||| |    |  |||||  ||| |||||    |||| | |  |||||  |
Db    428 TTCCGCAAGGTCACCAAGGACTTGGAGACGCGGGTTATCCGGTGGTTTGACTACCTGTGG 487

Qy    934 atcaacaagaagatgaccaacgaggtagccatcttacagcacttgcctgagcggctgcgg 993
          ||||||||||| |    | |||   |  |  |  ||  | | || ||   ||||  |
Db    488 GCCAACAAGAAGACGGTGGATGAGAAGGAGGTGCTCAAGAGCCTCCCAGACAAGCTGAAG 547

Qy    994 gcagaagtggctgtgtctgtgcacctgtccactctgagccgggtgcagatctttcagaac 1053
          || ||  | ||  |    |||||||||  ||| ||||    ||| |  |||||  ||| ||
Db    548 GCTGAGATCGCCATCAACGTGCACCTGGACACGCTGAAGAAGGTTCGCATCTTCCAGGAC 607

Qy   1054 tgtgaggccagcctgctgg 1072
          ||||| || || |||||||
Db    608 TGTGAAGCAAGGCTGCTGG 626
```

RESULT 6
BF190904
LOCUS BF190904 559 bp mRNA linear EST 02-NOV-2000
DEFINITION 237569 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BF190904
VERSION BF190904.1 GI:11074273
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 559)
AUTHORS Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W. and Keele,J.W.
TITLE Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine
JOURNAL Unpublished (2000)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -minmatch 12 options.
PCR PRimers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTTCCCAGTCACGACG
Plate: 65 row: J column: 10
Seq primer: ATTTAGGTGACACTATAG.
FEATURES Location/Qualifiers
source 1. .559
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 2PIG"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI; Library made from pooled tissue from testis, ovary, endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT 184 a 94 c 136 g 145 t
ORIGIN

Query Match 13.4%; Score 230.8; DB 10; Length 559;
Best Local Similarity 63.5%; Pred. No. 1.4e-45;
Matches 352; Conservative 0; Mismatches 202; Indels 0; Gaps 0;

```
Qy    919 tatcagcacctgcagatcaacaagaagatgaccaacgaggtagccatcttacagcacttg 978
          | | |  | |||  || ||||||| || |       | |||  ||   ||||  || |  |
Db      1 TTTGACTATCTGTGGACCAACAAAAAAACAGTAGATGAGAAAGAAGTCTTGAAGTATCTA 60

Qy    979 cctgagcggctgcgggcagaagtggctgtgtctgtgcacctgtccactctgagccgggtg 1038
          |||||    ||  | ||||||| | ||  |     || ||| |   |||  | |    ||||
Db     61 CCTGATAAACTAAGAGCAGAAATTGCCATTAACGTTCACTTAGACACATTAAAAAAGGTG 120

Qy   1039 cagatctttcagaactgtgaggccagcctgctggaggagctggtgctgaagctgcagccc 1098
          |  || |||    ||||||||||  | ||  ||| |||| ||||  ||||  | ||||||
Db    121 CGTATTTTTGCAGACTGTGAGGCTGGTCTCTTGGTGGAGTTGGTCTTGAAATTACAGCCC 180

Qy   1099 cagacctactcaccaggtgaatatgtatgccgcaaaggagacattggccaagagatgtac 1158
          |||  ||||   || || || ||  | |||   ||||| || ||||| | ||||||||||
Db    181 CAGGTCTACAGTCCTGGAGATTACATTTGCAAGAAAGGGGATATTGGACGAGAGATGTAC 240

Qy   1159 atcatccgagagggtcaactggccgtggtggcagatgatggtatcacacagtatgctgtg 1218
          || |||  ||| ||  |||| || |||||||||||||||| ||||| |||| ||  ||
Db    241 ATTATCAAAGAAGGCAAACTTGCTGTGGTGGCAGATGATGGAATCACCCAGTTTGTGGTA 300

Qy   1219 ctcggtgcagggctctactttggggagatcagcatcatcaacatcaaagggaacatgtct 1278
          |  |||  ||   |||||||| |||||||||| | ||||| ||||| |  |  ||
Db    301 TTGAGTGATGGCAGCTACTTTGGTGAGATCAGCATCCTTAACATTAAAGGCAGTAAAGCT 360

Qy   1279 gggaaccgccgcacagccaacatcaagagcctaggttattcagacctattctgcctgagc 1338
          || || ||  | |||||||| || || ||  | || |||||||| || ||||| ||
Db    361 GGCAATCGAAGAACAGCCAATATTAAAAGTATTGGCTATTCAGATCTGTTCTGTCTCTCG 420

Qy   1339 aaggaggacctgcgggaggtgctgagcgagtatccacaagcacagaccatcatggaggag 1398
          || ||  |||||   ||| |  || | |||||| ||| | ||  |    ||  | || |||
Db    421 AAAGATGACCTCATGGAAGCTCTAACCGAGTACCCAGATGCTAAATGTATGCTAGAAGAG 480

Qy   1399 aaaggacgtgagatcctgctgaaaatgaacaagttggacgtgaatgctgaggcagctgag 1458
          |||||    | ||| || |||||        | ||||  | |||  ||     ||||
Db    481 AAAGGGAAGCAAATCTTGATGAAAGATGGTCTGCTGGATATCAATATTGCAAATGCTGGA 540

Qy   1459 atcgccctgcagga 1472
          |  | ||   | ||
Db    541 AGTGACCCCAAAGA 554
```

RESULT 7
BF190899
LOCUS BF190899 560 bp mRNA linear EST 02-NOV-2000

```
DEFINITION  237563 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BF190899
VERSION     BF190899.1  GI:11074268
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 560)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 65  row: I  column: 10
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .560
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 2PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from testis, ovary,
                     endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT      183 a      95 c     136 g     146 t
ORIGIN
```

```
  Query Match           13.2%;  Score 227.4;  DB 10;  Length 560;
  Best Local Similarity  63.8%;  Pred. No. 9.6e-45;
  Matches  345;  Conservative    0;  Mismatches  196;  Indels    0;  Gaps    0;

Qy    919 tatcagcacctgcagatcaacaagaagatgaccaacgaggtagccatcttacagcacttg 978
          | | |  | |||  || |||||| || |      | |||  ||   ||||  || |  |
Db      1 TTTGACTATCTGTGGACCAACAAAAAAACAGTAGATGAGAAAGAAGTCTTGAAGTATCTA 60

Qy    979 cctgagcggctgcgggcagaagtggctgtgtctgtgcacctgtccactctgagccgggtg 1038
          |||||    ||  | |||||| | ||  |    || ||| |   |||  | |    ||||
Db     61 CCTGATAAACTAAGAGCAGAAATTGCCATTAACGTTCACTTAGACACATTAAAAAAGGTG 120

Qy   1039 cagatctttcagaactgtgaggccagcctgctggaggagctggtgctgaagctgcagccc 1098
          |  || |||    ||||||||||  | ||  ||| |||| |||| ||||  | ||||||
Db    121 CGTATTTTTGCAGACTGTGAGGCTGGTCTCTTGGTGGAGTTGGTCTTGAAATTACAGCCC 180

Qy   1099 cagacctactcaccaggtgaatatgtatgccgcaaaggagacattggccaagagatgtac 1158
          |||  ||||   || || || ||  | |||   ||||| || ||||| | ||||||||||
Db    181 CAGGTCTACAGTCCTGGAGATTACATTTGCAAGAAAGGGGATATTGGACGAGAGATGTAC 240

Qy   1159 atcatccgagagggtcaactggccgtggtggcagatgatggtatcacacagtatgctgtg 1218
          || |||  ||| ||  |||| || ||||||||||||||||| ||||| || | ||  ||
Db    241 ATTATCAAAGAAGGCAAACTTGCTGTGGTGGCAGATGATGGAATCACCCACTTTGTGGTA 300

Qy   1219 ctcggtgcagggctctactttggggagatcagcatcatcaacatcaaagggaacatgtct 1278
           |  |||  ||   |||||||| |||||||||||| | ||||| ||||| |  |   ||
Db    301 TTGAGTGATGGCAGCTACTTTGGTGAGATCAGCATCCTTAACATTAAAGGCAGTAAAGCT 360

Qy   1279 gggaaccgccgcacagccaacatcaagagcctaggttattcagacctattctgcctgagc 1338
          ||  | ||  | |||||||| || || ||  | || |||||||| || ||||| ||
Db    361 GGCTATCGAAGAACAGCCAATATTAAAAGTATTGGCTATTCAGATCTGTTCTGTCTCTCG 420

Qy   1339 aaggaggacctgcgggaggtgctgagcgagtatccacaagcacagaccatcatggaggag 1398
          || || |||||   ||| |  || | |||||| ||| | ||  |    ||  | || |||
Db    421 AAAGATGACCTCATGGAAGCTCTAACCGAGTACCCAGATGCTAAATGTATGCTAGAAGAG 480

Qy   1399 aaaggacgtgagatcctgctgaaaatgaacaagttggacgtgaatgctgaggcagctgag 1458
          |||||    | ||| || |||||        | ||||  | |||  ||     |||| |
Db    481 AAAGGGAAGCAAATCTTGATGAAAGATGGTCTGCTGGATATCAATATTGCAAATGCTGGG 540

Qy   1459 a 1459
          |
Db    541 A 541
```

```
RESULT   8
AW640105
LOCUS       AW640105                558 bp    mRNA    linear   EST 26-APR-2001
DEFINITION  bl91d05.w1 Blackshear/Soares normalized Xenopus egg library Xenopus
            laevis cDNA clone PBX0091D05 5', mRNA sequence.
ACCESSION   AW640105
VERSION     AW640105.1  GI:7397317
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 558)
  AUTHORS   Blackshear,P.J., Lai,W.S., Thorn,J.M., Kennington,E.A., Staffa,N.G.
            Jr., Moore,D.T., Bouffard,G.G., Beckstrom-Sternberg,S.M., Touchman
            ,J.W., Bonaldo,M.F. and Soares,M.B.
  TITLE     The NIEHS Xenopus maternal EST project: interim analysis of the
            first 13,879 ESTs from unfertilized eggs
  JOURNAL   Gene 267 (1), 71-87 (2001)
  MEDLINE   21211403
COMMENT     Contact: Perry J. Blackshear
            Office of Clinical Research and Laboratory of Signal Transduction
            National Institute of Environmental Health Sciences
            A2-05 NIEHS, 101 Alexander Drive, Research Triangle Park, NC 27709,
            USA
            Tel: 919 541-4899
            Fax: 919 541-4571
            Email: black009@niehs.nih.gov
            Clone is available through Research Genetics, Inc., 2130 Memorial
            Parkway, Huntsville, AL 35901
             phone 800-533-4363 ext.cdna, fax 256-536-9016 att:cdna, email
            cdna@resgen.com
             DNA Sequencing and analyses performed by National Institutes of
            Health Intramural Sequencing Center (NISC).
            PCR PRimers
            FORWARD: TGTAAAACGACGGCCAGT
            BACKWARD: CAGGAAACAGCTATGACC
            Plate: 0091  row: D  column: 05
            Seq primer: T7 primer.
FEATURES             Location/Qualifiers
     source          1. .558
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="PBX0091D05"
                     /clone_lib="Blackshear/Soares normalized Xenopus egg
                     library"
                     /sex="female"
                     /tissue_type="unfertilized egg"
                     /cell_type="unfertilized egg"
                     /dev_stage="unfertilized egg"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3-Pac; Site_1: EcoRI; Site_2: NotI;
                     PolyA-selected mRNA was prepared from unfertilized Xenopus
                     laevis eggs.  The library was constructed in the vector
                     pT7T3-Pac as described in Bonaldo, M.F., Lennon, G. and
                     Soares, M.B. 'Normalization and subtraction:  two
                     approaches to facilitate gene discovery',  Genome Research
```

6:791-806, 1996. The first strand synthesis used a NotI-dT18 primer; double stranded cDNAs were ligated to EcoRI adapters, digested with NotI, and directionally cloned into the NotI and EcoRI-digested pT7T3-Pac vector. The library contained approximately 7.2 X 10^5 recombinants, with average insert sizes of 1-1.5 kb."
BASE COUNT 191 a 97 c 136 g 133 t 1 others
ORIGIN

Query Match 12.8%; Score 221.6; DB 9; Length 558;
Best Local Similarity 64.6%; Pred. No. 2.5e-43;
Matches 329; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

```
Qy    914 actggtatcagcacctgcagatcaacaagaagatgaccaacgaggtagccatcttacagc 973
          ||  |  | |  ||||   ||  ||||| || |  |   | |||   |   | ||| ||
Db      6 ACGAGGTTGACTACCTCTGGACAAACAAAAAAACCATGGATGAGAAGGAGGTATTAAAGA 65

Qy    974 acttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctgtccactctgagcc 1033
          |||||| ||      |   ||| ||| | ||  |   ||| ||  |   |||| | |
Db     66 GCTTGCCAGACAAATTAAAGGCTGAAATTGCAATTAATGTACATTTAGACACTTTAAAAA 125

Qy   1034 gggtgcagatctttcagaactgtgaggccagcctgctggaggagctggtgctgaagctgc 1093
            ||  |||||||||  |||| |||||  |  ||||    || ||||| ||||| ||
Db    126 AAGTTAGGATCTTTCAAGACTGCGAGGCTGGATTGCTTATAGAACTGGTACTGAAACTAA 185

Qy   1094 agccccagacctactcaccaggtgaatatgtatgccgcaaaggagacattggccaagaga 1153
          ||||||   ||     || ||||| ||  | |||||  |||||||| |||||    || |
Db    186 GGCCCCAAGTCTTTAGCCCTGGTGACTACATTTGCCGAAAAGGAGATATTGGAAGGGAAA 245

Qy   1154 tgtacatcatccgagagggtcaactggccgtggtggcagatgatggtatcacacagtatg 1213
          ||||||||||  ||||||  |||| || || || || |||||||| ||||| |||| ||
Db    246 TGTACATCATCAAAGAGGGGAAACTAGCTGTAGTAGCTGATGATGGCATCACTCAGTTTG 305

Qy   1214 ctgtgctcggtgcagggctctactttggggagatcagcatcatcaacatcaaagggaaca 1273
            || ||| |||  ||    || ||||| || || |||||| | ||||||||||| | ||
Db    306 TGGTCCTCAGTGATGGCAGTTATTTTGGCGAAATTAGCATCTTAAACATCAAAGGAAGCA 365

Qy   1274 tgtctgggaaccgccgcacagccaacatcaagagcctaggttattcagacctattctgcc 1333
          ||| || |||||  | ||||||||||||||  ||  | || |||||  |||||  || ||
Db    366 AGTCAGGCAACCGGAGGACAGCCAACATCAGAAGTATTGGCTATTCCGACCTGTTTTGTT 425

Qy   1334 tgagcaaggaggacctgcgggaggtgctgagcgagtatccacaagcacagaccatcatgg 1393
          |    ||||| ||  ||  |||||  || |  || |||||  | ||  | ||  || | |
Db    426 TATCAAAGGATGATTTGATGGAGGCACTTACTGAATATCCTGATGCTAAAACAGTCTTAG 485

Qy   1394 aggagaaaggacgtgagatcctgctgaaa 1422
          | || ||||| ||  ||||||   |||||
Db    486 AAGAAAAAGGGCGACAGATCCNCATGAAA 514
```

RESULT 9
AZ503692/c
LOCUS AZ503692 621 bp DNA linear GSS 05-OCT-2000
DEFINITION 1M0343L05R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0343L05 R, DNA sequence.
ACCESSION AZ503692
VERSION AZ503692.1 GI:10685008
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 621)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0343 row: L column: 05
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 621.
FEATURES Location/Qualifiers
source 1. .621
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0343L05"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 143 a 166 c 123 g 189 t
ORIGIN

Query Match 12.6%; Score 217.8; DB 12; Length 621;
Best Local Similarity 62.8%; Pred. No. 2.2e-42;
Matches 355; Conservative 0; Mismatches 207; Indels 3; Gaps 1;

```
Qy    997 gaagtggctgtgtctgtgcacctgtccactctgagccgggtgcagatctttcagaactgt 1056
          ||| | ||  |    || || ||   |||  | |    ||| |  ||||||    |||||
Db    621 GAAATTGCCATCAACGTTCATCTAGACACGTTAAAAAAGGTTCGCATCTTTGCTGACTGT 562

Qy   1057 gaggccagcctgctggaggagctggtgctgaagctgcagccccagacctactcaccaggt 1116
          || ||  | ||| ||| |||| ||||| |||| || || ||||||    ||    || ||
Db    561 GAAGCTGGTCTGTTGGTGGAGTTGGTGTTGAAACTACAACCCCAGGTGTATAGTCCCGGA 502

Qy   1117 gaatatgtatgccgcaaaggagacattggccaagagatgtacatcatccgagagggtcaa 1176
          || ||  |||||   ||||| ||||| || |  ||||||||||||||   || ||  ||
Db    501 GATTACATATGCAAGAAAGGGGACATCGGGCGGGAGATGTACATCATCAAGGAAGGCAAA 442

Qy   1177 ctggccgtggtggcagatgatggtatcacacagtatgctgtgctcggtgcagggctctac 1236
          || || ||||||||||||||| || || ||||||| ||  ||| |  |||  ||    |||
Db    441 CTTGCTGTGGTGGCAGATGACGGAATTACACAGTTTGTGGTGTTGAGTGACGGCAGTTAC 382

Qy   1237 tttggggagatcagcatcatcaacatcaaagggaacatgtctgggaaccgccgcacagcc 1296
          ||||| ||||||||||   | || |||||||| | || | |||| |||||  | ||||||
Db    381 TTTGGTGAGATCAGCATTCTTAATATCAAAGGCAGCAAGGCTGGCAACCGAAGAACAGCC 322

Qy   1297 aacatcaagagcctaggttattcagacctattctgcctgagcaaggaggacctgcgggag 1356
          ||||| |||||| | || || || |||||  ||||||||   ||| || |||||   |||
Db    321 AACATTAAGAGCATCGGCTACTCGGACCTGTTCTGCCTCTCCAAAGATGACCTCATGGAA 262

Qy   1357 gtgctgagcgagtatccacaagcacagaccatcatggaggagaaaggacgtgagatcctg 1416
          |  || |  ||||| ||| | ||  | || ||  |||| ||||||||  |  ||||| |
```

```
Db    261 GCTCTTACAGAGTACCCAGATGCCAAAACTATGCTGGAAGAGAAAGGGAGGCAGATCTTA 202

Qy   1417 ctgaaaatgaacaagttggacgtgaatgctgaggcagctgagatcgccctgcaggaggcc 1476
          |||||        | ||||  | ||   ||         |  |  | ||   | ||
Db    201 ATGAAAGATGGTCTGCTGGATATAAACATTGCAAATATGGGCAGTGACCCTAAAGACCTG 142

Qy   1477 acagagtcccggctacgaggcctagaccagcagctggatgatctacagaccaagtttgct 1536
            ||||     | |      |  | ||   |    ||||    || || ||    |||||
Db    141 GAAGAG---AAGGTCACTCGAATGGAGGGGTCAGTGGACCTCCTGCAAACGCGATTTGCC 85

Qy   1537 cgcctcctggctgagctggagtcca 1561
          ||  || |||||||    ||||| |
Db     84 CGTATCTTGGCTGAATATGAGTCGA 60


RESULT 10
BF937274
LOCUS       BF937274                 567 bp    mRNA    linear   EST 13-FEB-2002
DEFINITION  fm62h03.y1 Zebrafish adult retina cDNA Danio rerio cDNA clone
            4200053 5' similar to SW:CNG1_BOVIN Q00194 CGMP-GATED CATION
            CHANNEL PROTEIN ;, mRNA sequence.
ACCESSION   BF937274
VERSION     BF937274.1  GI:12354594
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 567)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            Library constructed by: Chandra Tucker and Gregory Niemi DNA
            Sequencing by: Washington University Genome Sequencing Center Clone
            distribution: RessourcenZentrumPrimarDatenbank, Berlin, Germany
            (web address: www.rzpd.de)
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 420.
FEATURES             Location/Qualifiers
     source          1. .567
                     /organism="Danio rerio"
                     /strain="wild-type"
                     /db_xref="taxon:7955"
                     /clone="4200053"
                     /clone_lib="Zebrafish adult retina cDNA"
                     /sex="mixed"
                     /dev_stage="1-2 years"
                     /lab_host="E.Coli XL1-Blue MRF' (XL1-Blue MRF')"
                     /note="Vector: Lambda ZAP II (pBluescript SK-); Site_1:
                     EcoRI; Site_2: SalI; This Zebrafish library was
                     constructed by Dr. Susan E. Brockerhoff (email:
                     sbrocker@u.washington.edu) RZPD library number: 760"
BASE COUNT      156 a    133 c    143 g    134 t      1 others
ORIGIN


  Query Match             12.5%;  Score 215.6;  DB 10;  Length 567;
  Best Local Similarity   61.2%;  Pred. No. 7.2e-42;
  Matches  347;  Conservative    0;  Mismatches  220;  Indels    0;  Gaps    0;

Qy    607 tacccggaccccgcgcagcctggctttgagcgcctgcggcgccagtacctctatagcttt 666
          ||||||||||  |  | ||||  ||||  ||  ||   ||  | ||   ||| ||| |
Db      1 TACCCGGACCCTTCAGATCCTGAGTTTGGCCGATTGGTACGTAAATATGCCTACAGCATG 60

Qy    667 tacttctccacgctgatactgactacagtgggcgatacaccgccgccagccagggaagaa 726
          ||||  ||||| ||||  || |||||  | ||||| || |||||||| | |    |
Db     61 TACTGGTCCACCCTGACGCTCACTACCATCGGCGAAACGCCGCCGCCCGTCGAAAACTCT 120

Qy    727 gagtacctcttcatggtgggcgacttcctgctggccgtcatgggtttcgccaccatcatg 786
          |||||| ||||| | ||   |||||| ||  | | |||  ||  |||||| || ||| |
Db    121 GAGTACTTCTTCGTAGTCACCGACTTTCTAGTAGGCGTGTTGATTTTCGCAACAATCGTC 180

Qy    787 ggtagcatgagctctgtcatctacaacatgaacactgcagatgcggctttctacccagat 846
          |||| | |  ||||  | |||   ||    ||| | |||   || | |||| |  |
Db    181 GGTAACGTCGGCTCAATGATCACAAATGCCAACGCAGCACGAGCCGATTTCCAAGCTCGC 240

Qy    847 catgcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggagcggcga 906
            |     | ||  | |||||||   | | | | |||| |    |  |||||  ||
Db    241 ATCGATGCTATTAAACAATACATGAGCTTTCGGAAAGTCACCAAGGATTTGGAGAAACGG 300

Qy    907 gttattgactggtatcagcacctgcagatcaacaagaagatgaccaacgaggtagccatc 966
          |||||| | |||| | |  |  ||  || ||| || |||  |  | | |||  ||   |
Db    301 GTTATTAAATGGTTTGATTATTTGTGGACCAATAAAAAGGCGGTCGATGAGAGAGAAGTA 360

Qy    967 ttacagcacttgcctgagcggctgcgggcagaagtggctgtgtctgtgcacctgtccact 1026
            |  || |  | || ||   |||| |||| ||  | ||  |    || || |||   ||
Db    361 CTGAAGTATCTTCCCGACAAGCTGAGGGCGGAGATAGCCATAAACGTCCATCTGGATACG 420

Qy   1027 ctgagccgggtgcagatctttcagaactgtgaggccagcctgctggaggagctggtgctg 1086
          || |     ||||  |||| |  | |||||| ||| | ||  ||| ||||||||| |
Db    421 CTAAAAAAAGTGCGTATCTNTGCGGACTGTGAAGCCGGGCTTTTGGTGGAGCTGGTGTTA 480

Qy   1087 aagctgcagccccagacctactcaccaggtgaatatgtatgccgcaaaggagacattggc 1146
          |||||||||||  |||   |||   || || || ||| |||||   || || |||||  |
Db    481 AAGCTGCAGCCGCAGGTTTACAGCCCTGGAGACTATATATGCAAAAACGGCGACATCCGG 540

Qy   1147 caagagatgtacatcatccgagagggt 1173
          |  ||||||||||||||   ||| |||
Db    541 CGCGAGATGTACATCATTAAAGACGGT 567


RESULT 11
CNS02XOL
LOCUS       CNS02XOL                 819 bp    DNA     linear   GSS 15-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            179G05 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL218622
VERSION     AL218622.1  GI:7877441
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 819)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 819)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 819)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
```

```
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .819
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="179G05"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG179AD03SP1-end :
                     PUC-Ori"
BASE COUNT      217 a    236 c    180 g    186 t
ORIGIN
```

```
  Query Match             11.8%;  Score 204.4;  DB 12;  Length 819;
  Best Local Similarity   54.5%;  Pred. No. 4.7e-39;
  Matches  409;  Conservative    0;  Mismatches  341;  Indels    0;  Gaps    0;

Qy    249 catggtggtgcgcttccacacaggattcttggaacagggcatcctggtggtggacaaggg 308
          | |  ||   |  || | |    | | |||||| || ||| | ||||||   |||
Db     67 CCTAATGTCACCTTTGCCCCGCAGTTACTTGGAGCAAGGCCTTCTGGTGAAAGACGCCAA 126

Qy    309 taggatctcgagtcgctacgttcgcacctggagtttcttcttggacctggcttccctgat 368
          |   |      |  |||       ||       |||   |  ||| ||  |||| ||||
Db    127 GAAACTGAGAGATAAATACCGAACAACACCCCAGTTCAAGTACGACATGATTTCCATGAT 186

Qy    369 gcccacagatgtggtctacgtgcggctgggcccgcacacacccaccctgaggctgaaccg 428
          |||||  ||  || | |   |       |||    |||  ||     |  |  | |||||
Db    187 ACCCACCGACCTGCTGTTTCTCAACTACGGCTACAACAATCCGGAGTTCCGCTTCAACCG 246

Qy    429 ctttctccgcgcgccccgcctcttcgaggccttcgaccgcacagagacccgcacagctta 488
          | ||  |       |  | || || ||    || ||  | || || ||  | ||    |
Db    247 CCTTTGCAAAATCTCAAGACTTTTTGAATTTTTTGAAAGGACTGAAACTAGAACCAGCTT 306

Qy    489 cccaaatgcctttcgcattgccaagctgatgctttacatttttgtcgtcatccattggaa 548
          ||| ||    || || ||   ||| ||  |||| |||||   | ||| |||||| |||||
Db    307 CCCCAACATGTTCCGTATCAGCAACCTAGTGCTCTACATCCTCGTCATCATCCACTGGAA 366

Qy    549 cagctgcctatactttgccctatcccggtacctgggcttcgggcgtgacgcatgggtgta 608
          |  |||||| | ||| ||||| |||     | | ||||| ||  ||||| | ||||||||
Db    367 CGCCTGCCTGTTCTTCGCCCTCTCCAAAACCATCGGCTTTGGCAGTGACACCTGGGTGTA 426

Qy    609 cccggaccccgcgcagcctggctttgagcgcctgcggcgccagtacctctatagctttta 668
          ||  ||  |   || || |    ||  ||||||    |  ||||| ||||     | ||
Db    427 TCCCAACATCAGCCACCCGGAGCATGGCCGCCTGGCCAGGAAGTACATCTACTCGCTGTA 486

Qy    669 cttctccacgctgatactgactacagtgggcgatacaccgccgccagccagggaagaaga 728
          ||  ||||||||||  || || ||  | || || || ||  | || | || |||    ||
Db    487 CTGGTCCACGCTGACCCTCACCACCATCGGAGAGACGCCTGCACCCGTCAAGGATATCGA 546

Qy    729 gtacctcttcatggtgggcgacttcctgctggccgtcatgggtttcgccaccatcatggg 788
          |||||  ||| |  |    ||||||||    | |||  ||   || || |  ||| | ||
Db    547 ATACCTTTTCGTTATTTCGGACTTCCTCACCGGCGTGCTGATCTTTGCTAGTATCGTCGG 606

Qy    789 tagcatgagctctgtcatctacaacatgaacactgcagatgcggctttctacccagatca 848
          | | |  |  |  | ||||  |||||||| || |    || |  ||  |  || |
Db    607 AAACGTCGGTGCCATGATCTCAAACATGAATGCTTCTCGGGCCGAGTTTCAGGCAAAAAT 666

Qy    849 tgcactggtgaagaagtacatgaagctgcagcacgtcaaccgcaagctggagcggcgagt 908
          |      | ||| |||||||| || | |   | |||  |     | ||||||    | ||
Db    667 CGACTCCATAAAGCAGTACATGCAGTTCCGAAAGGTCTCCAAAGAACTGGAGGCCAGGGT 726

Qy    909 tattgactggtatcagcacctgcagatcaacaagaagatgaccaacgaggtagccatctt 968
          ||  | ||||   |  |||||  ||   | |||||||    | | |||   |   || |
Db    727 CATCAAGTGGTTCGACTACCTGTGGACGGAGAAGAAGACCTGCGATGAGAAGGAAGTCCT 786

Qy    969 acagcacttgcctgagcggctgcgggcaga 998
          || || | || ||   |||    || ||
Db    787 GAAGAACCTCCCGGACAAGCTCAAAGCCGA 816
```

```
RESULT  12
BG293093
LOCUS       BG293093                882 bp    mRNA    linear   EST 21-FEB-2001
DEFINITION  602389921F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4501732 5',
            mRNA sequence.
ACCESSION   BG293093
VERSION     BG293093.1  GI:13052518
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 882)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM10369  row: j  column: 05
           High quality sequence stop: 685.
FEATURES             Location/Qualifiers
     source          1. .882
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4501732"
                     /clone_lib="NIH_MGC_94"
                     /tissue_type="retina"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 3.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      261 a    178 c    203 g    240 t
ORIGIN
```

```
  Query Match             11.1%;  Score 192.4;  DB 10;  Length 882;
  Best Local Similarity   57.9%;  Pred. No. 4e-36;
  Matches  360;  Conservative    0;  Mismatches  261;  Indels    1;  Gaps    1;

Qy    467 gcacagagacccgcacagcttacccaaatgcctttcgcattgccaagctgatgctttaca 526
          | || || ||  | ||    ||||||||   |||| | ||| | || ||  || | ||||
Db      3 GGACGGAAACAAGGACCAACTACCCAAACATCTTTAGGATTTCGAACCTTGTGATGTACA 62

Qy    527 tttttgtcgtcatccattggaacagctgcctatactttgccctatcccggtacctgggct 586
          |  |  || |||||| ||||||   ||  | || | | || | ||        |||  |
Db     63 TCGTCATCATCATCCACTGGAACGCTTGTGTGTATTATTCCATCTCAAAAGCTATGGATT 122

Qy    587 tcgggcgtgacgcatgggtgtacccggaccccgcgcagcctggctttgagcgcctgcggc 646
          | |||  |||| |||||| ||||| ||        | ||||  || |  ||  |
Db    123 T-GGGAATGACACATGGGTCTACCCTGATGTTAATGATCCTGAATTCGGCCGTTTAGCTA 181

Qy    647 gccagtacctctatagcttttacttctccacgctgatactgactacagtgggcgatacac 706
          |  | ||  |||| ||| |||| |  || ||  |||   |||| ||  | || || | |
Db    182 GAAAATATGTCTACAGCCTTTATTGGTCTACTTTGACTTTGACAACCATTGGTGAAACTC 241

Qy    707 cgccgccagccagggaagaagagtacctcttcatggtgggcgacttcctgctggccgtca 766
          | || || |    |||    |||||| |||| ||||||  |||||| | | |  ||
Db    242 CACCTCCCGTGCTGGATTCTGAGTACATCTTTGTGGTGGTAGACTTCTTAATTGGAGTTT 301

Qy    767 tgggtttcgccaccatcatgggtagcatgagctctgtcatctacaacatgaacactgcag 826
          |   |||||||||||| | |||| |||  | ||  | || | ||| |||||  | ||
Db    302 TAATTTTCGCCACCATTGTCGGTAACATAGGTTCCATGATTTCCAATATGAATGCAGCCC 361

Qy    827 atgcggctttctacccagatcatgcactggtgaagaagtacatgaagctgcagcacgtca 886
```

```
Db    362 GGGCAGAATTTCAATCAAGAGTTGATGCTATCAAACAGTACATGAATTTTCGAAATGTAA 421

Qy    887 accgcaagctggagcggcgagttattgactggtatcagcacctgcagatcaacaagaaga 946
Db    422 GCAAAGATATGGAAAAGAGGGTTATTAAATGGTTTGACTACCTGTGGACCAACAAAAAAA 481

Qy    947 tgaccaacgaggtagccatcttacagcacttgcctgagcggctgcgggcagaagtggctg 1006
Db    482 CAGTGGATGAGAGAGAAGTCCTGAGATACCTCCCTGACAAACTAAGGGCAGAAATTGCCA 541

Qy   1007 tgtctgtgcacctgtccactctgagccgggtgcagatctttcagaactgtgaggccagcc 1066
Db    542 TCAACGTTCATCTAGACACGTTAAAAAAGGTTCGCATCTTTGCTGACTGTGAAGCTGGTC 601

Qy   1067 tgctggaggagctggtgctgaa 1088
Db    602 TGTGGTGGGAGTTGGTGTTGGA 623


RESULT 13
CNS042AH/c
LOCUS       CNS042AH                923 bp    DNA     linear   GSS 18-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            077A21 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL271250
VERSION     AL271250.1  GI:7993216
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 923)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 923)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 923)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .923
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="077A21"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG077AA11LP1~end : T7"
BASE COUNT      256 a    208 c    264 g    194 t      1 others
ORIGIN


  Query Match           10.4%;  Score 179.6;  DB 12;  Length 923;
  Best Local Similarity  56.9%;  Pred. No. 5.4e-33;
  Matches  354;  Conservative    0;  Mismatches  259;  Indels    9;  Gaps    1;

Qy    269 caggattcttggaacagggcatcctggtggtggacaagggtaggatctcgagtcgctacg 328
Db    668 CAGGTTACCTGGAACAAGGCCTGCTGGTGAAAGACGAGAAGCTCCTCAGGGACGGCTACG 609

Qy    329 ttcgcacctggagtttcttcttggacctggcttccctgatgcccacagatgtggtctacg 388
Db    608 TTGCCAGCTTTCAGTTTCGTCTTGATGTCGTTTCCATGCTTCCCACTGACATCTTGTACT 549

Qy    389 tgcggctgggcccgcacacacccaccctgaggctgaaccgctttctccgcgcgccccgcc 448
Db    548 TGTGCTTTGGCCTCGACTACCCAGAGATCCGAATCAACAAGCTTTTGAGAATCGGCCGGA 489

Qy    449 tcttcgaggccttcgaccgcacagagacccgcacagcttacccaaatgcctttcgcattg 508
Db    488 TGATGGAGTTCTTCACCAGGACCGAGACTAAAACCAACTACCCCAACATCTTCCGCATTG 429

Qy    509 ccaagctgatgctttacatttttgtcgtcatccattggaacagctgcctatactttgccc 568
Db    428 CAAACCTGATTATGTACATCCTCATTATCATCCACTGGAACGCTTGCTTTTACTTCTCCT 369

Qy    569 tatcccggtacctgggcttcgggcgtgacgcatgggtgtaccc---------ggaccccg 619
Db    368 TTTCCAAATCTATCGGTTTTGGCGCTGACGAATGGGTCTACCCAGCTCTGGACGATCCGG 309

Qy    620 cgcagcctggctttgagcgcctgcggcgccagtacctctatagcttttacttctccacgc 679
Db    308 ACGAGCCTGCTTTCGGGGAGGCCATGAGGAAGTATTCTTTCAGCCTGTACTGGTCTACCC 249

Qy    680 tgatactgactacagtgggcgatacaccgccgccagccagggaagaagagtacctcttca 739
Db    248 TGACCCTCACCACCATTGGAGAAACCCCCCCACCGGCCCTGGACTCGGAATTCCTCTTTC 189

Qy    740 tggtgggcgacttcctgctggccgtcatgggtttcgccaccatcatgggtagcatgagct 799
Db    188 ATGTGATTGATTTCCTAGTTGGCGTCCTGATTTTTGCCACCATTGTAGGAAACATCGCCA 129

Qy    800 ctgtcatctacaacatgaacactgcagatgcggctttctacccagatcatgcactggtga 859
Db    128 CCATGATCTCAAACATGAACGCCGCCCAGGCTCAGTTCCAGGCTCGGATCGACAACATGA 69

Qy    860 agaagtacatgaagctgcagca 881
Db     68 AACACTACATGCAGGTACAGCA 47


RESULT 14
BE491188/c
LOCUS       BE491188                580 bp    mRNA    linear   EST 31-JUL-2000
DEFINITION  db43c03.x1 Blackshear/Soares normalized Xenopus egg library Xenopus
            laevis cDNA clone IMAGE:3301252 3' similar to gb:S42457 CGMP-GATED
            CATION CHANNEL PROTEIN (HUMAN); gb:U19717 Mus musculus cGMP-gated
            cation channel mRNA, clone mIMCD-K2 (MOUSE);, mRNA sequence.
ACCESSION   BE491188
VERSION     BE491188.1  GI:9610649
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 580)
  AUTHORS   Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.
            , Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
            ,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     WashU Xenopus EST project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Sandy Clifton, Ph.D.
            WashU Xenopus EST project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
```

```
            Library constructed by Bento Soares and M. Fatima Bonaldo
            (University of Iowa). DNA Sequencing by: Washington University
            Genome Sequencing Center
             Clone distribution: Xenopus clones from this library are available
            through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
            Seq primer: -40UP from Gibco
            High quality sequence stop: 453.
FEATURES              Location/Qualifiers
     source           1. .580
                      /organism="Xenopus laevis"
                      /db_xref="taxon:8355"
                      /clone="IMAGE:3301252"
                      /clone_lib="Blackshear/Soares normalized Xenopus egg
                      library"
                      /sex="female"
                      /tissue_type="unfertilized egg"
                      /cell_type="unfertilized egg"
                      /dev_stage="unfertilized egg"
                      /lab_host="DH10B"
                      /note="Vector: pT7T3-Pac; Site_1: EcoRI; Site_2: NotI;
                      PolyA-selected mRNA was prepared from unfertilized Xenopus
                      laevis eggs.  The library was constructed in the vector
                      pT7T3-Pac as described in Bonaldo, M.F., Lennon, G. and
                      Soares, M.B. 'Normalization and subtraction:  two
                      approaches to facilitate gene discovery',  Genome Research
                      6:791-806, 1996.  The first strand synthesis used a
                      NotI-dT18 primer; double stranded cDNAs were ligated to
                      EcoRI adapters, digested with NotI, and directionally
                      cloned into the NotI and EcoRI-digested pT7T3-Pac vector.
                      The library contained approximately 7.2 X 10^5
                      recombinants, with average insert sizes of 1-1.5 kb."
BASE COUNT      130 a    148 c     95 g    207 t
ORIGIN


  Query Match             9.7%;  Score 167.4;  DB 10;  Length 580;
  Best Local Similarity   66.5%;  Pred. No. 4e-30;
  Matches  240;  Conservative    0;  Mismatches  121;  Indels    0;  Gaps    0;

Qy   1112 caggtgaatatgtatgccgcaaaggagacattggccaagagatgtacatcatccgagagg 1171
           | |||||  ||  | |||||  ||||||||  |||||    ||  ||||||||||||  |||||
Db    567 CTGGTGACTACATTTGCCGAAAAGGAGATATTGGAAGGGAAATGTACATCATCAAAGAGG 508

Qy   1172 gtcaactggccgtggtggcagatgatggtatcacacagtatgctgtgctcggtgcagggc 1231
           |  ||||  ||  ||  ||  ||  ||||||||  |||||  ||||  ||  ||  |||  |||  ||
Db    507 GGAAACTAGCTGTAGTAGCTGATGATGGCATCACTCAGTTTGTGGTCCTCAGTGATGGCA 448

Qy   1232 tctactttggggagatcagcatcatcaacatcaaagggaacatgtctgggaaccgccgca 1291
             ||  |||||  ||  ||  ||||||  |  ||||||||||  |  ||  |||  ||  |||||  |  |
Db    447 GTTATTTTGGCGAAATTAGCATCTTAAACATCAAAGGAAGCAAGTCAGGCAACCGGAGGA 388

Qy   1292 cagccaacatcaagagcctaggttattcagacctattctgcctgagcaaggaggacctgc 1351
           ||||||||||||  ||  |  ||  |||||  |||||  ||  ||  |    |||||  ||  ||
Db    387 CAGCCAACATCAGAAGTATTGGCTATTCCGACCTGTTTTGTTTATCAAAGGATGATTTGA 328

Qy   1352 gggaggtgctgagcgagtatccacaagcacagaccatcatggaggagaaaggacgtgaga 1411
           |||||  ||  |  ||  |||||  |  ||  |  ||  ||  |  ||  ||  |||||  ||  |||
Db    327 TGGAGGCACTTACTGAATATCCTGATGCTAAAACAGTCTTAGAAGAAAAAGGGCGACAGA 268

Qy   1412 tcctgctgaaaatgaacaagttggacgtgaatgctgaggcagctgagatcgccctgcagg 1471
           ||||  |||||          |  ||  |  |  ||  |    |||  |    |  ||  |  |||
Db    267 TCCTCATGAAAGATGGTTTAATAGATGAAGAAGCAGCCAAAGCAGGTGCAGACCCGAAGG 208

Qy   1472 a 1472
           |
Db    207 A 207


RESULT  15
W84514
LOCUS       W84514                   502 bp    mRNA    linear   EST 17-OCT-1996
DEFINITION  zd90g10.r1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone
            IMAGE:356802 5' similar to gb:S42457 CGMP-GATED CATION CHANNEL
            PROTEIN (HUMAN);, mRNA sequence.
ACCESSION   W84514
VERSION     W84514.1  GI:1395695
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 502)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1035   Std Error: 0.00
            Seq primer: ETPrimer
            High quality sequence stop: 413.
FEATURES              Location/Qualifiers
     source           1. .502
                      /organism="Homo sapiens"
                      /db_xref="GDB:1273346"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:356802"
                      /clone_lib="Soares_fetal_heart_NbHH19W"
                      /sex="unknown"
                      /dev_stage="19 weeks"
                      /lab_host="DH10B (ampicillin resistant)"
                      /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
                      modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                      strand cDNA was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCATCTTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was size selected, ligated to Eco RI
                      adapters (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of a modified pT7T3 vector
                      (Pharmacia). Library went through one round of
                      normalization to a Cot = 5. Library constructed by
                      M.Fatima Bonaldo. This library was constructed from the
                      same fetus as the fetal lung library, Soares fetal lung
                      NbHL19W."
BASE COUNT      141 a     94 c    131 g    120 t     16 others
ORIGIN


  Query Match             8.7%;  Score 150.8;  DB 10;  Length 502;
  Best Local Similarity   61.7%;  Pred. No. 4.1e-26;
  Matches  258;  Conservative    0;  Mismatches  157;  Indels    3;  Gaps    1;

Qy   1144 ggccaagagatgtacatcatccgagagggtcaactggccgtggtggcagatgatggtatc 1203
           || | ||||||||||||| |||   || ||  |||| || ||||||||||||||||||  ||
Db      2 GGACGAGAGATGTACATTATCAAGGAAGGCAAACTCGCTGTGGTGGCAGATGATGGAGTC 61

Qy   1204 acacagtatgctgtgctcggtgcagggctctactttggggagatcagcatcatcaacatc 1263
           || |||| ||  ||  |  | |  ||   ||||||  ||  ||||||||||  | |||||
Db     62 ACTCAGTTTGTGGTATTGAGCGATGGCAGCTACTTCGGTGAGATCAGCATTCTTAACATT 121

Qy   1264 aaagggaacatgtctgggaaccgccgcacagccaacatcaagagcctaggttattcagac 1323
           |||||||  ||   ||||  ||  ||  |  ||  |||||  ||  ||  ||   |  ||  ||  ||||||
Db    122 AAAGGGAGCAAAGCTGGCAATCGAAGAACGGCCAATATTAAAAGTATTGGCTACTCAGAC 181

Qy   1324 ctattctgcctgagcaaggaggacctgcgggaggtgctgagcgagtatccacaagcacag 1383
           ||  |||||  ||    ||  ||  |||||   |||  |  ||  |  |||||  |||  |  ||  |
Db    182 CTGTTCTGTCTCTCAAAAGATGACCTCATGGAAGCTCTAACTGAGTACCCAGATGCCAAA 241

Qy   1384 accatcatggaggagaaaggacgtgagatcctgctgaaaatgaacaagttggacgtgaat 1443
```

```
          || ||  |||| ||||||||     ||||  |  |||||          ||||  | ||
Db    242 ACTATGCTGGAAGAGAAAGGGAAGCAGATTTTAATGAAAGATGGTCTACTGGATCTAAAC 301

Qy   1444 gctgaggcagctgagatcgccctgcaggaggccacagagtcccggctacgaggcctagac 1503
            ||     ||||  |  |  |   | ||      ||||    | |     |  | ||
Db    302 ATTGCAAATGCTGGCAGTGATCCTAAAGATCTTGAAGAG---AAGGTTACTCGAATGGAG 358

Qy   1504 cagcagctggatgatctacagaccaagtttgctcgcctcctggctgagctggagtcca 1561
            |    | ||    || || |||| |||||| ||  || ||||||||   |||||||
Db    359 GGGTCAGTAGACCTCCTGCAAACCAGGTTTGCCCGAATCTTGGCTGAGTATGAGTCCA 416
```

Search completed: June 20, 2002, 12:16:58
Job time: 11371 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 20, 2002, 14:39:52 ; Search time 48.82 Seconds
(without alignments)
1131.736 Million cell updates/sec

Title: US-09-927-267-1
Perfect score: 2989
Sequence: 1 MSQDTKVKTTESSPPAPSKA.....EGTSKDEEGRASQEGPPGPE 575

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2811	94.0	575	2	I59327	olfactory cyclic n
2	1579	52.8	706	2	A55251	cyclic nucleotide-
3	1568	52.5	663	2	S11521	cAMP-gated channel
4	1562.5	52.3	735	2	I50630	alpha subunit of c
5	1562	52.3	732	2	S35691	cyclic nucleotide-
6	1558	52.1	664	2	S11517	cyclic nucleotide-
7	1544.5	51.7	690	1	S07103	cGMP-gated ion cha
8	1542	51.6	695	2	S74179	cyclic nucleotide-
9	1538.5	51.5	686	1	A44842	cGMP-gated ion cha
10	1532.5	51.3	691	2	JC6509	rod cyclic nucleot
11	1529	51.2	645	2	I50680	alpha subunit of r
12	1526.5	51.1	690	2	A42161	cGMP-gated cation
13	1525.5	51.0	688	2	B42161	cGMP-gated cation
14	1436.5	48.1	682	1	JH0560	cyclic nucleotide-
15	1235	41.3	665	2	S52072	DmCNGC protein - f
16	1121.5	37.5	772	2	S28292	hypothetical prote
17	858	28.7	261	2	I78560	cyclic nucleotide-
18	847	28.3	261	2	I78559	cyclic nucleotide-
19	788.5	26.4	800	2	T19627	hypothetical prote
20	738	24.7	673	2	T20936	hypothetical prote
21	716	24.0	909	2	S32538	cGMP-gated cation
22	647	21.6	644	2	T33125	hypothetical prote
23	642	21.5	611	2	T20935	hypothetical prote
24	511.5	17.1	191	2	S74158	cGMP-gated cation
25	491.5	16.4	189	2	S74159	cGMP-gated cation
26	470.5	15.7	767	2	T21969	hypothetical prote
27	387.5	13.0	962	2	I53197	potassium channel
28	382.5	12.8	1087	2	T31100	probable potassium
29	382.5	12.8	1284	2	T13168	probable potassium
30	377.5	12.6	1174	2	A40853	potassium channel
31	377	12.6	989	2	I48912	potassium channel
32	375	12.5	1159	2	I38465	probable potassium
33	370	12.4	1102	2	T17367	potassium channel
34	365	12.2	934	2	T42394	potassium channel
35	362.5	12.1	514	2	T19579	hypothetical prote
36	362	12.1	1017	2	T31354	probable potassium
37	339.5	11.4	828	2	T52046	potassium channel
38	310.5	10.4	662	2	T04461	potassium channel
39	307.5	10.3	807	2	T12177	potassium channel
40	300	10.0	688	2	S55349	potassium channel
41	295.5	9.9	716	2	T51354	cyclic nucleotide-
42	293	9.8	787	2	S68699	potassium channel
43	285.5	9.6	845	2	T07052	probable potassium
44	269.5	9.0	887	2	T03939	potassium channel
45	268.5	9.0	838	2	S23606	potassium channel

ALIGNMENTS

RESULT 1
I59327
olfactory cyclic nucleotide gated cation channel - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 31-Mar-2000
C;Accession: I59327; I58165
R;Bradley, J.; Li, J.; Davidson, N.; Lester, H.A.; Zinn, K.
Proc. Natl. Acad. Sci. U.S.A. 91, 8890-8894, 1994
A;Title: Heteromeric olfactory cyclic nucleotide-gated channels: A new subunit that c
A;Reference number: I59327; MUID:94377458
A;Accession: I59327
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-575 <RES>
A;Cross-references: EMBL:U12623; NID:g538128; PIDN:AAA21464.1; PID:g538129
R;Liman, E.R.; Buck, L.B.
Neuron 13, 611-621, 1994
A;Title: A second subunit of the olfactory cyclic nucleotide-gated channel confers hi
A;Reference number: I58165; MUID:95000663
A;Accession: I58165
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-575 <RE2>
A;Cross-references: EMBL:U12425; NID:g548083; PIDN:AAA64748.1; PID:g548084
C;Superfamily: cyclic nucleotide-gated channel; cAMP receptor protein cyclic nucleoti
F;348-472/Domain: cAMP receptor protein cyclic nucleotide-binding domain homology <CA

Query Match 94.0%; Score 2811; DB 2; Length 575;
Best Local Similarity 93.4%; Pred. No. 5.5e-188;
Matches 537; Conservative 16; Mismatches 22; Indels 0; Gaps 0;

```
Qy       1 MSQDTKVKTTESSPPAPSKARKLLPVLDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPD 60
           |||| ||||||:||||:|||| ||||||||||||||||||||:||||||:||||||||
Db       1 MSQDGKVKTTESTPPAPTKARKWLPVLDPSGDYYYWWLNTMVFPIMYNLIIVVCRACFPD 60

Qy      61 LQHGYLVAWLVLDYTSDLLYLLDMVVRFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLA 120
           ||| ||||| |||||||||||||: |||||||||||||||||| |:|||||||| ||||
Db      61 LQHSYLVAWFVLDYTSDLLYLLDIGVRFHTGFLEQGILVVDKGMIASRYVRTWSFLLDLA 120

Qy     121 SLMPTDVVYVRLGPHTPTLRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVI 180
           ||:|||  ||:|||| |||||||||| ||||||||||||||||||||||||||||||||
Db     121 SLVPTDAAYVQLGPHIPTLRLNRFLRVPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVI 180

Qy     181 HWNSCLYFALSRYLGFGRDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAR 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||| | |
Db     181 HWNSCLYFALSRYLGFGRDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPLPDR 240

Qy     241 EEEYLFMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLE 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||::||
Db     241 EEEYLFMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNKRLE 300
```

```
Qy     301 RRVIDWYQHLQINKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEEL 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 RRVIDWYQHLQINKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEEL 360

Qy     361 VLKLQPQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIIN 420
           ||||||||||||||||||||:||||||||||||||||||:||||||||||||||||||||
Db     361 VLKLQPQTYSPGEYVCRKGDIGREMYIIREGQLAVVADDGVTQYAVLGAGLYFGEISIIN 420

Qy     421 IKGNMSGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLDV 480
           |||||||||||||||||||||||||||||||||||||||| :||||||||||||||||||
Db     421 IKGNMSGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQAQAVMEEKGREILLKMNKLDV 480

Qy     481 NAEAAEIALQEATESRLRGLDQQLDDLQTKFARLLAELESSALKIAYRIERLEWQTREWP 540
           ||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db     481 NAEAAEIALQEATESRLKGLDQQLDDLQTKFARLLAELESSALKIAYRIERLEWQTREWP 540

Qy     541 MPEDLAEADDEGEPEEGTSKDEEGRASQEGPPGPE 575
           ||||: ||||| || |||||| ||:| | || | |
Db     541 MPEDMGEADDEAEPGEGTSKDGEGKAGQAGPSGIE 575


RESULT   2
A55251
cyclic nucleotide-gated Ca++ channel protein CNG-3 - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 13-Aug-1999
C;Accession: A55251; S43976
R;Biel, M.; Zong, X.; Distler, M.; Bosse, E.; Klugbauer, N.; Murakami, M.; Flockerzi, V.
Proc. Natl. Acad. Sci. U.S.A. 91, 3505-3509, 1994
A;Title: Another member of the cyclic nucleotide-gated channel family, expressed in test
A;Reference number: A55251; MUID:94224768
A;Accession: A55251
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-706 <BIE>
A;Cross-references: GB:X76485; NID:g488728; PIDN:CAA54023.1; PID:g488729
A;Experimental source: kidney
R;Weyand, I.; Godde, M.; Frings, S.; Weiner, J.; Mueller, F.; Altenhofen,.W.; Hatt, H.;
Nature 368, 859-863, 1994
A;Title: Cloning and functional expression of a cyclic-nucleotide-gated channel from mam
A;Reference number: S43976; MUID:94211295
A;Accession: S43976
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-706 <WEY>
A;Cross-references: GB:X89600; NID:g908823; PIDN:CAA61759.1; PID:g908824
A;Experimental source: testis
C;Genetics:
A;Gene: CNG3
C;Superfamily: cyclic nucleotide-gated channel; cAMP receptor protein cyclic nucleotide-
C;Keywords: cGMP binding; ion channel; ion transport; membrane protein
F;501-625/Domain: cAMP receptor protein cyclic nucleotide-binding domain homology <CAP>


  Query Match            52.8%;  Score 1579;  DB 2;  Length 706;
  Best Local Similarity  52.7%;  Pred. No. 4.7e-102;
  Matches  297;  Conservative  111;  Mismatches  138;  Indels   18;  Gaps    3;

Qy       3 QDTKVKTTESSP-------PAPSKARKLLPVLDPSGDYYYWWLNTMVFPVMYNLIILVCR 55
           :| | |  |          |   : :|   |:||| : || ||  :  || ||  :||||
Db     149 KDDKAKKEEKEKKEEKKENPKKEEKKKDSVVMDPSSNMYYHWLTVIAVPVFYNWCLLVCR 208

Qy      56 ACFPDLQHGYLVAWLVLDYTSDLLYLLDMVVRFHTGFLEQGILVVDKGRISSRYVRTWSF 115
           ||| :||  :|: |||||||::|:|| :||:||  |||||||::|:|  |:   | :|  |
Db     209 ACFDELQSEHLMLWLVLDYSADILYGMDMLVRARTGFLEQGLMVMDASRLWKHYTQTLHF 268

Qy     116 FLDLASLMPTDVVYVRLGPHTPTLRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLY 175
            ||: ||:|||: | :|| : | || || |:  |||| |||||||| ||| |||  |:||
Db     269 KLDVLSLVPTDLAYFKLGMNYPELRFNRLLKLARLFEFFDRTETRTNYPNMFRIGNLVLY 328

Qy     176 IFVVIHWNSCLYFALSRYLGFGRDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDT 235
           | ::||||:|:|||:|:::||| |:|||: : | : || |:|:|| |:||| |||:|:|
Db     329 ILIIIHWNACIYFAISKFIGFGTDSWVYPNVSNPEYGRLSRKYIYSLYWSTLTLTTIGET 388

Qy     236 PPPAREEEYLFMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHV 295
           ||| ::|||||:| |||: |: ||||:|:: |:| ||| : | |      :|:||: : |
Db     389 PPPVKDEEYLFVVIDFLVGVLIFATIVGNVGSMISNMNASRAEFQAKIDSIKQYMQFRKV 448

Qy     296 NRKLERRVIDWYQHLQINKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEAS 355
           : || ||| |: :|  ||| :|  :|: ||::|:||:|::||| || :|:|||:|||
Db     449 TKDLETRVIRWFDYLWANKKTVDEKEVLKSLPDKLKAEIAINVHLDTLRKVRIFQDCEAG 508

Qy     356 LLEELVLKLQPQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGE 415
           || ||||||:|  :|||:|:|:|||||:|||||:||:|||||:|||||: ||| | ||||
Db     509 LLVELVLKLRPAVFSPGDYICKKGDIGREMYIIKEGKLAVVAEDGITQFVVLGDGSYFGE 568

Qy     416 ISIINIKGNMSGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKM 475
           |||:||||: ||||||||||:|:|||||||||||:|| | |:|||:|:  :|||||:||:|
Db     569 ISILNIKGSKSGNRRTANIRSIGYSDLFCLSKDDLMEALTEYPEAKKALEEKGRQILMKD 628

Qy     476 NKLD-----VNAEAAEIALQEATESRLRGLDQQLDDLQTKFARLLAELESSALKIAYRIE 530
           | :|       |:  :|      | ::  |:  || |||:|||||||  :: :|:  |:
Db     629 NLIDEELAKAGADPKDI------EEKVEHLETSLDSLQTRFARLLAEYNATQMKVKQRLS 682

Qy     531 RLEWQTREWPMPEDLAEADDEGEP 554
           :|| | :    |:  |   :  :|
Db     683 QLESQVKMGLPPDGDAPQTEASQP 706


RESULT   3
S11521
cAMP-gated channel protein - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 21-Jan-2000
C;Accession: S11521
R;Ludwig, J.; Margalit, T.; Eismann, E.; Lancet, D.; Kaupp, U.B.
FEBS Lett. 270, 24-29, 1990
A;Title: Primary structure of cAMP-gated channel from bovine olfactory epithelium.
A;Reference number: S11521; MUID:91032022
A;Accession: S11521
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-663 <LUD>
A;Cross-references: GB:X55010; NID:g287743; PIDN:CAA38754.1; PID:g287744
C;Superfamily: cyclic nucleotide-gated channel; cAMP receptor protein cyclic nucleoti
F;454-578/Domain: cAMP receptor protein cyclic nucleotide-binding domain homology <CA


  Query Match            52.5%;  Score 1568;  DB 2;  Length 663;
  Best Local Similarity  54.2%;  Pred. No. 2.5e-101;
  Matches  292;  Conservative  108;  Mismatches  133;  Indels    6;  Gaps    2;

Qy      18 SKARKLLPVLDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHGYLVAWLVLDYTSD 77
           :| :  | ||||:||:|| ||  :  ||:||  :|| |||| ||| || : |||||| ||
Db     124 TKKKFELFVLDPAGDWYYRWLFLIALPVLYNWCLLVARACFSDLQKGYYIVWLVLDYVSD 183

Qy      78 LLYLLDMVVRFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLASLMPTDVVYVRLGPHTP 137
           ::|: |: :|  |||||||:|| |  ::   |: |  | ||:||::|||::|  :| | |
Db     184 VVYIADLFIRLRTGFLEQGLLVKDTKKLRDNYIHTMQFKLDVASIIPTDLIYFAVGIHNP 243

Qy     138 TLRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSCLYFALSRYLGFG 197
            :| || |   |:|| |||||||:||| |||: |:||| ::||||:|:|:|:|: :|||
Db     244 EVRFNRLLHFARMFEFFDRTETRTSYPNIFRISNLILYILIIIHWNACIYYAISKSIGFG 303

Qy     198 RDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAREEEYLFMVGDFLLAVMG 257
           | ||||:   | :  | |:|:|  |:||| |||:|:|||| ::|||||:: |||: |:
Db     304 VDTWVYPNITDPEYGYLSREYIYCLYWSTLTLTTIGETPPPVKDEEYLFVIFDFLIGVLI 363

Qy     258 FATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVIDWYQHLQINKKMT 317
           ||||:|:: |:| |||   | |      || ||: : |::::| :|| |: :|  |||
Db     364 FATIVGNVGSMISNMNATRAEFQAKIDAVKHYMQFRKVSKEMEAKVIRWFDYLWTNKKSV 423

Qy     318 NEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQPQTYSPGEYVCR 377
           :|  :|::|| :||||:|::|||||| :|:|||:||| || ||||||:|| :|||:|:||
Db     424 DEREVLKNLPAKLRAEIAINVHLSTLKKVRIFQDCEAGLLVELVLKLRPQVFSPGDYICR 483
```

```
Qy    378 KGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIINIKGNMSGNRRTANIKSL 437
          |||||:|||||:||:|||||||||:||||:| ||  |||||||:||||:  ||||||||:|| 
Db    484 KGDIGKEMYIIKEGKLAVVADDGVTQYALLSAGSCFGEISILNIKGSKMGNRRTANIRSL 543

Qy    438 GYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLDVNAEAA--EIALQEATES 495
          ||||||||||:|| | ::||| |: ::||:|||||:|   || |  ||  |: :|| 
Db    544 GYSDLFCLSKDDLMEAVTEYPDAKRVLEERGREILMKEGLLDENEVAASMEVDVQE---- 599

Qy    496 RLRGLDQQLDDLQTKFARLLAELESSALKIAYRIERLEWQTREWPMPEDLAEADDEGEP 554
          :|  |:  :| | |:|||||||   :  |:  ||  || : ::    : |::  :  || 
Db    600 KLEQLETNMDTLYTRFARLLAEYTGAQQKLKQRITVLETKMKQNNEDDSLSDGMNSPEP 658


RESULT   4
I50630
alpha subunit of cone photoreceptor CNG-channel - chicken
C;Species: Gallus gallus (chicken)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 13-Aug-1999
C;Accession: I50630
R;Bonigk, W.; Altenhofen, W.; Muller, F.; Dose, A.; Illing, M.; Molday, R.S.; Kaupp, U.B
Neuron 10, 865-877, 1993
A;Title: Rod and cone photoreceptor cells express distinct genes for cGMP-gated channels
A;Reference number: I50630; MUID:93264082
A;Accession: I50630
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-735 <BON>
A;Cross-references: EMBL:X89598; NID:g908850; PIDN:CAA61757.1; PID:g908851
C;Superfamily: cyclic nucleotide-gated channel; cAMP receptor protein cyclic nucleotide-
F;524-648/Domain: cAMP receptor protein cyclic nucleotide-binding domain homology <CAP>


  Query Match             52.3%;  Score 1562.5;  DB 2;  Length 735;
  Best Local Similarity   53.5%;  Pred. No. 7e-101;
  Matches  299;  Conservative  108;  Mismatches  149;  Indels    3;  Gaps    3;

Qy      3 QDTKVKTTESSPPAPSKARKLLPVLDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQ 62
          :| | |  :       : :| : |:||| : || ||  :  || ||  :|:|||||| :|| 
Db    180 KDDK-KDDKKDDKKKEEQKKEVFVIDPSSNMYYNWLTIIAAPVFYNWCMLICRACFDELQ 238

Qy     63 HGYLVAWLVLDYTSDLLYLLDMVVRFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLASL 122
            :: `|| ||| ||:::|: || ||| |||||||:|| |: ::   | :|  | ||: || 
Db    239 IDHIKLWLFLDYCSDIIYVFDMFVRFRTGFLEQGLLVKDEKKLRDHYTQTVQFKLDVLSL 298

Qy    123 MPTDVVYVRLGPHTPTLRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHW 182
          :|||: |::|| : | || || ||  |||| |||||||| ||| |||  |:||| ::|||
Db    299 LPTDLAYLKLGLNYPELRFNRLLRIARLFEFFDRTETRTNYPNMFRIGNLVLYILIIIHW 358

Qy    183 NSCLYFALSRYLGFGRDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAREE 242
          |:|:|||:|: :||| |:|||: : | : || |:|:|| |:||| |||:|:|||| ::|
Db    359 NACIYFAISKVIGFGTDSWVYPNVSIPEYGRLSRKYIYSLYWSTLTLTTIGETPPPVKDE 418

Qy    243 EYLFMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERR 302
          |||||:| |||: |: ||||:|:: |:| ||| : | |      :|:||  : | : || | 
Db    419 EYLFVVIDFLVGVLIFATIVGNVGSMISNMNASRAEFQAKVDSIKQYMHFRKVTKDLEAR 478

Qy    303 VIDWYQHLQINKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVL 362
          || |: :|  |||  :|  :|::||::|:||:|::||| || :|:|||:||| || |||| 
Db    479 VIKWFDYLWTNKKTVDEKEVLKNLPDKLKAEIAINVHLDTLKKVRIFQDCEAGLLIELVL 538

Qy    363 KLQPQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIINIK 422
          ||:|  :|||:|:|:|||||:|||||:||:|||||||||||: ||  | |||||||:||| 
Db    539 KLKPTVFSPGDYICKKGDIGREMYIIKEGKLAVVADDGITQFVVLSDGSYFGEISILNIK 598

Qy    423 GNMSGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLDVNA 482
          |: |||||||||:|:|||||||||:|| | |:|||:|:  :|||||:||:| | :|  | 
Db    599 GSKSGNRRTANIRSIGYSDLFCLSKDDLMEALTEYPEAKKALEEKGRQILMKDNLIDEEA 658

Qy    483 EAAEIALQEATESRLRGLDQQLDDLQTKFARLLAELESSALKIAYRIERLEWQTREWPMP 542
            |  |  :  | ::  |:  || |||:|||||||  ||  |:  |: |:| : :::  
Db    659 AKAG-ADPKDLEEKIDRLETALDTLQTRFARLLAEYSSSQQKVKQRLARVETRVKKYG-S 716

Qy    543 EDLAEADDEGEPEEGTSKD 561
```

```
          |:  : | |  |   ||
Db    717 GSLSVGEPEPEKPEEQKKD 735


RESULT   5
S35691
cyclic nucleotide-gated channel protein - rabbit
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 13-Aug-1999
C;Accession: S35691
R;Biel, M.; Altenhofen, W.; Hullin, R.; Ludwig, J.; Freichel, M.; Flockerzi, V.; Dasc
FEBS Lett. 329, 134-138, 1993
A;Title: Primary structure and functional expression of a cyclic nucleotide-gated cha
A;Reference number: S35691; MUID:93359035
A;Accession: S35691
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-732 <BIE>
A;Cross-references: EMBL:X59668; NID:g433959; PIDN:CAA42201.1; PID:g433960
C;Superfamily: cyclic nucleotide-gated channel; cAMP receptor protein cyclic nucleoti
F;522-646/Domain: cAMP receptor protein cyclic nucleotide-binding domain homology <CA


  Query Match             52.3%;  Score 1562;  DB 2;  Length 732;
  Best Local Similarity   54.2%;  Pred. No. 7.6e-101;
  Matches  292;  Conservative  108;  Mismatches  133;  Indels    6;  Gaps    2;

Qy     18 SKARKLLPVLDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHGYLVAWLVLDYTSD 77
          :| :  | ||||:||:|| ||  :  ||:||  :|| |||| ||| || : |||||| ||
Db    192 TKKKFELFVLDPAGDWYYRWLFVIAMPVLYNWCLLVARACFSDLQRGYFLVWLVLDYFSD 251

Qy     78 LLYLLDMVVRFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLASLMPTDVVYVRLGPHTP 137
          ::|: |: :|  |||||||:|| |  ::   |: |  | ||:||::|||::|  :| | |
Db    252 VVYIADLFIRLRTGFLEQGLLVKDPKKLRDNYIHTLQFKLDVASIIPTDLIYFAVGIHNP 311

Qy    138 TLRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSCLYFALSRYLGFG 197
            || || |   |:|| ||||||||:||| |||: |:||| |:||||:|:|:|:|: :|||
Db    312 ELRFNRLLHFARMFEFFDRTETRTSYPNIFRISNLVLYILVIIHWNACIYYAISKSIGFG 371

Qy    198 RDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAREEEYLFMVGDFLLAVMG 257
          | ||||:   | :  | |:|:|  |:||| |||:|:|||| ::|||||:: |||: |:
Db    372 VDTWVYPNITDPEYGYLAREYIYCLYWSTLTLTTIGETPPPVKDEEYLFVIFDFLIGVLI 431

Qy    258 FATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVIDWYQHLQINKKMT 317
          ||||:|:: |:| |||   | |       || ||: : |::::| :|| |: :|  |||
Db    432 FATIVGNVGSMISNMNATRAEFQAKIDAVKHYMQFRKVSKEMEAKVIKWFDYLWTNKKTV 491

Qy    318 NEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQPQTYSPGEYVCR 377
          :|  :|::|| :||||:|::|||||| :|:|||:||| || ||||||:|| :|||:|:||
Db    492 DEREVLKNLPAKLRAEIAINVHLSTLKKVRIFQDCEAGLLVELVLKLRPQVFSPGDYICR 551

Qy    378 KGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIINIKGNMSGNRRTANIKSL 437
          |||||:|||||:||:|||||||||:||||:| ||  |||||||:||||:  ||||||||:|| 
Db    552 KGDIGKEMYIIKEGKLAVVADDGVTQYALLSAGSCFGEISILNIKGSKMGNRRTANIRSL 611

Qy    438 GYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLDVNAEAA--EIALQEATES 495
          ||||||||||:|| | ::||| |: ::||:|||||:|   || |  ||  |: :|| 
Db    612 GYSDLFCLSKDDLMEAVTEYPDAKKVLEERGREILMKEGLLDENEVAASMEVDVQE---- 667

Qy    496 RLRGLDQQLDDLQTKFARLLAELESSALKIAYRIERLEWQTREWPMPEDLAEADDEGEP 554
          :|: |:  :: | |:| |||||   :  |:  ||  || : ::    : |::  :  || 
Db    668 KLKQLETNMETLYTRFGRLLAEYTGAQQKLKQRITVLEVKMKQNTEDDYLSDGMNSPEP 726


RESULT   6
S11517
cyclic nucleotide-activated channel protein - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 18-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 13-Aug-1999
C;Accession: S11517
R;Dhallan, R.S.; Yau, K.W.; Schrader, K.A.; Reed, R.R.
Nature 347, 184-187, 1990
A;Title: Primary structure and functional expression of a cyclic nucleotide-activated
```

```
A;Reference number: S11517; MUID:90370115
A;Accession: S11517
A;Molecule type: mRNA
A;Residues: 1-664 <DHA>
A;Cross-references: GB:X55519; NID:g56791; PIDN:CAA39135.1; PID:g56792
C;Superfamily: cyclic nucleotide-gated channel; cAMP receptor protein cyclic nucleotide-
C;Keywords: transmembrane protein
F;456-580/Domain: cAMP receptor protein cyclic nucleotide-binding domain homology <CAP>


  Query Match           52.1%;  Score 1558;  DB 2;  Length 664;
  Best Local Similarity   54.2%;  Pred. No. 1.3e-100;
  Matches  292;  Conservative  106;  Mismatches  135;  Indels    6;  Gaps    2;

Qy      18 SKARKLLPVLDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHGYLVAWLVLDYTSD 77
           :| :  | ||||:||:|| ||  :  ||:||  :|| |||| ||| | | |||||| ||
Db     126 TKKKFELFVLDPAGDWYYRWLFVIAMPVLYNWCLLVARACFSDLQRNYFVVWLVLDYFSD 185

Qy      78 LLYLLDMVVRFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLASLMPTDVVYVRLGPHTP 137
            :|: |:::|  |||||||:|| |  ::   |: |  | ||:||::|||::|  :| |:|
Db     186 TVYIADLIIRLRTGFLEQGLLVKDPKKLRDNYIHTLQFKLDVASIIPTDLIYFAVGIHSP 245

Qy     138 TLRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSCLYFALSRYLGFG 197
           :| || |   |:|| ||||||||:||| |||: |:||| |:||||:|:|: :|: :|||
Db     246 EVRFNRLLHFARMFEFFDRTETRTSYPNIFRISNLVLYILVIIHWNACIYYVISKSIGFG 305

Qy     198 RDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAREEEYLFMVGDFLLAVMG 257
           | ||||:   | :  | |:|:|  |:||| |||:|:|||| ::|||||:: |||: |:
Db     306 VDTWVYPNITDPEYGYLAREYIYCLYWSTLTLTTIGETPPPVKDEEYLFVIFDFLIGVLI 365

Qy     258 FATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVIDWYQHLQINKKMT 317
           ||||:|:: |:| |||   | |       || ||: : |:: :| :|| |: :|  |||
Db     366 FATIVGNVGSMISNMNATRAEFQAKIDAVKHYMQFRKVSKDMEAKVIKWFDYLWTNKKTV 425

Qy     318 NEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQPQTYSPGEYVCR 377
           :|  :|::|| :|||||:|::|||||| :|:|||:||| || ||||||:|| :|||:|:||
Db     426 DEREVLKNLPAKLRAEIAINVHLSTLKKVRIFQDCEAGLLVELVLKLRPQVFSPGDYICR 485

Qy     378 KGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIINIKGNMSGNRRTANIKSL 437
           |||||:|||||:||:|||||||||:||||:| ||  |||||||:|||:  |||||||:||
Db     486 KGDIGKEMYIIKEGKLAVVADDGVTQYALLSAGSCFGEISILNIKGSKMGNRRTANIRSL 545

Qy     438 GYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLDVNAEAA--EIALQEATES 495
           ||||||||||:|| | ::||| |: ::||:|||||:|   || |  ||  |: :||
Db     546 GYSDLFCLSKDDLMEAVTEYPDAKKVLEERGREILMKEGLLDENEVAASMEVDVQE---- 601

Qy     496 RLRGLDQQLDDLQTKFARLLAELESSALKIAYRIERLEWQTREWPMPEDLAEADDEGEP 554
           :|  |:  :| | |:|||||||   :  |:  ||  || : ::    : |::  :  ||
Db     602 KLEQLETNMDTLYTRFARLLAEYTGAQQKLKQRITVLETKMKQNHEDDYLSDGINTPEP 660


RESULT   7
S07103
cGMP-gated ion channel protein - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: S07103
R;Kaupp, U.B.; Niidome, T.; Tanabe, T.; Terada, S.; Boenigk, W.; Stuehmer, W.; Cook, N.J
Nature 342, 762-766, 1989
A;Title: Primary structure and functional expression from complementary DNA of the rod p
A;Reference number: S07103; MUID:90098076
A;Accession: S07103
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-690 <KAU>
A;Cross-references: GB:X51604; NID:g203; PIDN:CAA35947.1; PID:g204
A;Note: part of this sequence was confirmed by protein sequencing
C;Superfamily: cyclic nucleotide-gated channel; cAMP receptor protein cyclic nucleotide-
C;Keywords: cGMP binding; eye; ion channel; ion transport; retina; transmembrane protein
F;477-601/Domain: cAMP receptor protein cyclic nucleotide-binding domain homology <CAP>


  Query Match           51.7%;  Score 1544.5;  DB 1;  Length 690;
  Best Local Similarity   54.0%;  Pred. No. 1.2e-99;
  Matches  285;  Conservative  109;  Mismatches  133;  Indels    1;  Gaps    1;

Qy       6 KVKTTESSPPAPSKARKLLPVLDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHGY 65
           |  |    |      : :| : |:|||| || ||  :  |||||  ::: |||| :|| |
Db     135 KKKKEEKGKDKKEEEKKEVVVIDPSGNTYYNWLFCITLPVMYNWTMIIARACFDELQSDY 194

Qy      66 LVAWLVLDYTSDLLYLLDMVVRFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLASLMPT 125
           |  ||  || ||::|||||  ||  ||:||||:|| :: ::  :|  |: | ||: |::||
Db     195 LEYWLAFDYLSDVVYLLDMFVRTRTGYLEQGLLVKEERKLIDKYKSTFQFKLDVLSVIPT 254

Qy     126 DVVYVRLGPHTPTLRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSC 185
           |::|:: | : | :|||| ||  |:|| | |||||| ||| |||: |::|| ::||||:|
Db     255 DLLYIKFGWNYPEIRLNRLLRISRMFEFFQRTETRTNYPNIFRISNLVMYIIIIIHWNAC 314

Qy     186 LYFALSRYLGFGRDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAREEEYL 245
           :||::|: :||| | |||||   | | || |:|:|| |:||| |||:|:|||| |: ||
Db     315 VYFSISKAIGFGNDTWVYPDVNDPDFGRLARKYVYSLYWSTLTLTTIGETPPPVRDSEYF 374

Qy     246 FMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVID 305
           |:| |||: |: ||||:|:: |:| ||| | | |      :|:||  ::|:: :|:|||
Db     375 FVVADFLIGVLIFATIVGNIGSMISNMNAARAEFQARIDAIKQYMHFRNVSKDMEKRVIK 434

Qy     306 WYQHLQINKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQ 365
           |: :|  |||  :|  :|:::||::||||:|::||| || :|:|| :||| || |||||||
Db     435 WFDYLWTNKKTVDEREVLKYLPDKLRAEIAINVHLDTLKKVRIFADCEAGLLVELVLKLQ 494

Qy     366 PQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIINIKGNM 425
           || ||||:|:|:|||||:|||||:||:|||||||||||: ||  | |||||||:||||:
Db     495 PQVYSPGDYICKKGDIGREMYIIKEGKLAVVADDGITQFVVLSDGSYFGEISILNIKGSK 554

Qy     426 SGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLDVNAEAA 485
           :|||||||||:|||||||||||:|| | |:||| |: ::|||::||:|   ||:|   |
Db     555 AGNRRTANIKSIGYSDLFCLSKDDLMEALTEYPDAKGMLEEKGKQILMKDGLLDINIANA 614

Qy     486 EIALQEATESRLRGLDQQLDDLQTKFARLLAELESSALKIAYRIERLE 533
           ::  | ::  ::  :| |||:|||:||| ||   |:  |: ::|
Db     615 GSDPKD-LEEKVTRMESSVDLLQTRFARILAEYESMQQKLKQRLTKVE 661


RESULT   8
S74179
cyclic nucleotide-gated channel protein - human
C;Species: Homo sapiens (man)
C;Date: 14-Apr-1998 #sequence_revision 24-Apr-1998 #text_change 16-Jul-1999
C;Accession: S74179
R;Yu, W.P.; Grunwald, M.E.; Yau, K.W.
FEBS Lett. 393, 211-215, 1996
A;Title: Molecular cloning, functional expression and chromosomal localization of a h
A;Reference number: S74179; MUID:96409310
A;Accession: S74179
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-695 <YUW>
A;Experimental source: retina
C;Genetics:
A;Map position: 2
C;Superfamily: cyclic nucleotide-gated channel; cAMP receptor protein cyclic nucleoti
C;Keywords: ion channel; ion transport; membrane protein
F;482-607/Domain: cAMP receptor protein cyclic nucleotide-binding domain homology <CA


  Query Match           51.6%;  Score 1542;  DB 2;  Length 695;
  Best Local Similarity   52.7%;  Pred. No. 1.7e-99;
  Matches  291;  Conservative  111;  Mismatches  140;  Indels   10;  Gaps    3;

Qy       9 TTESSPPAPSKARKLLPVLDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHGYLVA 68
           |: ::      :|   |:||| : || ||  :  || ||  :|:|||||| :|| ||:
Db     143 TSNNTEEEKKTKKKDAIVVDPSSNLYYRWLTAIALPVFYNWYLLICRACFDELQSEYLML 202

Qy      69 WLVLDYTSDLLYLLDMVVRFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLASLMPTDVV 128
           ||||||::|:||:||::||  |||||||::| |  |:   |  |  | ||: ||:|||:
Db     203 WLVLDYSADVLYVLDVLVRARTGFLEQGLMVSDTNRLWQHYKTTTQFKLDVLSLVPTDLA 262
```

```
Qy    129 YVRLGPHTPTLRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSCLYF 188
          |:::| : | :| || |:  |||| |||||||| ||| |||  |:||| ::|||||:|:|| 
Db    263 YLKVGTNYPEVRFNRLLKFSRLFEFFDRTETRTNYPNMFRIGNLVLYILIIIHWNACIYF 322

Qy    189 ALSRYLGFGRDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAREEEYLFMV 248
          |:|:::||| |:|||| : |   || |:|:|| |:||| |||:|:|||| ::|||||:| 
Db    323 AISKFIGFGTDSWVYPNISIPEHGRLSRKYIYSLYWSTLTLTTIGETPPPVKDEEYLFVV 382

Qy    249 GDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVIDWYQ 308
          |||: |: |||||:|:: |:| ||| : | |      :|:||: : | : || ||| |: 
Db    383 VDFLVGVLIFATIVGNVGSMISNMNASRAEFQAKIDSIKQYMQFRKVTKDLETRVIRWFD 442

Qy    309 HLQINKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQPQT 368
          :|  |||  :|  :|: ||::|:||:|::|||  || :|:|||:||| || ||||||:| 
Db    443 YLWANKKTVDEKEVLKSLPDKLKAEIAINVHLDTLKKVRIFQDCEAGLLVELVLKLRPTV 502

Qy    369 YSPGEYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIINIKGNMSGN 428
          :|||:|:|:|||||:||||| ||:||||||||:|| ||  | |||||||:|||||: ||| 
Db    503 FSPGDYICKKGDIGKEMYIINEGKLAVVADDGVTQFVVLSDGSYFGEISILNIKGSKSGN 562

Qy    429 RRTANIKSLGYSDLFCLSKEDLREVLSEYP-QAQTIMEEKGREILLKMNKLDVNAEAAEI 487
          ||||||:|:|||||||||||:|| | |:||| ||:  :|||||:||:| | :|     | 
Db    563 RRTANIRSIGYSDLFCLSKDDLMEALTEYPGQAKKALEEKGRQILMKDNLIDEELARAG- 621

Qy    488 ALQEATESRLRGLDQQLDDLQTKFARLLAELESSALKIAYRIERLEWQTR--------EW 539
          |  :  | ::  |   || |||:|||||||  :: :|:  |: :|| | :        : 
Db    622 ADPKDLEEKVEQLGSSLDTLQTRFARLLAEYNATQMKMKQRLSQLESQVKGGGDKPLADG 681

Qy    540 PMPEDLAEADDE 551
           :| |  : :|: 
Db    682 EVPGDATKTEDK 693


RESULT   9
A44842
cGMP-gated ion channel protein - human
N;Alternate names: rod photoreceptor cGMP-gated channel
C;Species: Homo sapiens (man)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A44842
R;Dhallan, R.S.; Macke, J.P.; Eddy, R.L.; Shows, T.B.; Reed, R.R.; Yau, K.W.; Nathans, J
J. Neurosci. 12, 3248-3256, 1992
A;Title: Human rod photoreceptor cGMP-gated channel: amino acid sequence, gene structure
A;Reference number: A44842; MUID:92356211
A;Accession: A44842
A;Molecule type: mRNA
A;Residues: 1-686 <DHA>
A;Cross-references: GB:S42457; NID:g252853; PIDN:AAB22778.1; PID:g252854
A;Experimental source: retina
A;Note: sequence extracted from NCBI backbone (NCBIN:110250, NCBIP:110251)
A;Note: intron positions were determined from genomic sequence
C;Genetics:
A;Gene: GDB:CNCG1; CNCG
A;Cross-references: GDB:127557; OMIM:123825
A;Map position: 4pl4-4q13
A;Introns: 36/2; 75/2; 95/2; 110/2; 146/2; 182/2; 218/2
C;Superfamily: cyclic nucleotide-gated channel; cAMP receptor protein cyclic nucleotide-
C;Keywords: cGMP binding; eye; ion channel; ion transport; retina; transmembrane protein
F;475-599/Domain: cAMP receptor protein cyclic nucleotide-binding domain homology <CAP>


  Query Match             51.5%;  Score 1538.5;  DB 1;  Length 686;
  Best Local Similarity   54.0%;  Pred. No. 3e-99;
  Matches  285;  Conservative  113;  Mismatches  129;  Indels    1;  Gaps    1;

Qy      6 KVKTTESSPPAPSKARKLLPVLDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHGY 65
          |  |  | |     : :| : |:|||||: || ||  :  |||||  ::: |||| :||  | 
Db    133 KKKKEEKSKDKKEEEKKEVVVIDPSGNTYYNWLFCITLPVMYNWTMVIARACFDELQSDY 192

Qy     66 LVAWLVLDYTSDLLYLLDMVVRFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLASLMPT 125
          |  ||:||| ||::||:|| ||  ||:||||:|| :: :: ::|     | ||: ||:|| 
Db    193 LEYWLILDYVSDIVYLIDMFVRTRTGYLEQGLLVKEELKLINKYKSNLQFKLDVLSLIPT 252

Qy    126 DVVYVRLGPHTPTLRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSC 185
          |::| :|| : | :|||| ||  |:|| | |||||| ||| |||: |::|| ::||||:| 
Db    253 DLLYFKLGWNYPEIRLNRLLRFSRMFEFFQRTETRTNYPNIFRISNLVMYIVIIIHWNAC 312

Qy    186 LYFALSRYLGFGRDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAREEEYL 245
          :::::|: :||| | |||||    | | || |:|:|| |:||| |||:|:|||| |: ||: 
Db    313 VFYSISKAIGFGNDTWVYPDINDPEFGRLARKYVYSLYWSTLTLTTIGETPPPVRDSEYV 372

Qy    246 FMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVID 305
          |:| |||: |: ||||:|:: |:| ||| | | |      :|:||  ::|:: :|:||| 
Db    373 FVVVDFLIGVLIFATIVGNIGSMISNMNAARAEFQARIDAIKQYMHFRNVSKDMEKRVIK 432

Qy    306 WYQHLQINKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQ 365
          |: :|  |||  :|  :|::||::||||:|::||| || :|:|| :||| || |||||||  
Db    433 WFDYLWTNKKTVDEKEVLKYLPDKLRAEIAINVHLDTLKKVRIFADCEAGLLVELVLKLQ 492

Qy    366 PQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIINIKGNM 425
          || ||||:|:|:|||||:|||||:||:|||||||||:|| ||  | |||||||:|||||: 
Db    493 PQVYSPGDYICKKGDIGREMYIIKEGKLAVVADDGVTQFVVLSDGSYFGEISILNIKGSK 552

Qy    426 SGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLDVNAEAA 485
          :|||||||||:|||||||||||:|| | |:||| |:|::||||::||:|   ||:|   | 
Db    553 AGNRRTANIKSIGYSDLFCLSKDDLMEALTEYPDAKTMLEEKGKQILMKDGLLDLNIANA 612

Qy    486 EIALQEATESRLRGLDQQLDDLQTKFARLLAELESSALKIAYRIERLE 533
          .      ::  | ::  ::  :| |||:|||:||| ||   |:  |: ::| 
Db    613 GSDPKD-LEEKVTRMEGSVDLLQTRFARILAEYESMQQKLKQRLTKVE 659


RESULT  10
JC6509
rod cyclic nucleotide-gated cation channel protein alpha-chain - dog
C;Species: Canis lupus familiaris (dog)
C;Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 21-Jul-2000
C;Accession: JC6509
R;Veske, A.; Nilsson, S.E.G.; Gal, A.
Gene 202, 115-119, 1997
A;Title: Characterization of canine rod photoreceptor cGMP-gated cation channel alpha
A;Reference number: JC6509; MUID:98087425
A;Accession: JC6509
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-691 <VES>
A;Cross-references: EMBL:X99914; NID:g1488571; PIDN:CAA68186.1; PID:g1488572
C;Superfamily: cyclic nucleotide-gated channel; cAMP receptor protein cyclic nucleoti
F;478-602/Domain: cAMP receptor protein cyclic nucleotide-binding domain homology <CA


  Query Match             51.3%;  Score 1532.5;  DB 2;  Length 691;
  Best Local Similarity   53.6%;  Pred. No. 7.9e-99;
  Matches  283;  Conservative  110;  Mismatches  134;  Indels    1;  Gaps    1;

Qy      6 KVKTTESSPPAPSKARKLLPVLDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHGY 65
          |    |       : :| : |:|||:|: || ||  :  |||||  ::: |||| :||  | 
Db    136 KKNKEEKGKDKKEEEKKEVMVIDPAGNMYYNWLFCITLPVMYNWTMVIARACFDELQSDY 195

Qy     66 LVAWLVLDYTSDLLYLLDMVVRFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLASLMPT 125
          |  |:: || ||::|||| ||  ||:||||:|| :: ::  :|     | ||  |::|| 
Db    196 LEYWIIFDYLSDIVYLLDMFVRTRTGYLEQGLLVREEAKLIEKYKSNLQFKLDFLSVIPT 255

Qy    126 DVVYVRLGPHTPTLRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSC 185
          |::| :|| : | :|||| ||  |:|| | |||||| ||| |||: |::|| ::||||:| 
Db    256 DLLYFKLGWNYPEIRLNRLLRISRMFEFFQRTETRTNYPNIFRISNLVMYIVIIIHWNAC 315

Qy    186 LYFALSRYLGFGRDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAREEEYL 245
          :||::|: :||| | |||||    | | || |:|:|| |:||| |||:|:|||| |: ||: 
Db    316 VYFSISKAIGFGNDTWVYPDVNDPEFGRLARKYVYSLYWSTLTLTTIGETPPPVRDSEYV 375

Qy    246 FMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVID 305
          |:| |||: |: ||||:|:: |:| ||| | | |      :|:||  ::|:: :|:||| 
Db    376 FVVVDFLIGVLIFATIVGNIGSMISNMNAARAEFQARIDAIKQYMHFRNVSKDMEKRVIK 435
```

```
Qy    306 WYQHLQINKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQ 365
          |: :|  |||  :|  :|::||::|||||:|::||| || :|:|| :||| || |||||||
Db    436 WFDYLWTNKKTVDEKEVLKYLPDKLRAEIAINVHLDTLKKVRIFADCEAGLLVELVLKLQ 495

Qy    366 PQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIINIKGNM 425
          || ||||:|:|:|||||:|||||:||:||||||||||||: ||  | |||||||:||||:
Db    496 PQVYSPGDYICKKGDIGREMYIIKEGKLAVVADDGITQFVVLSDGSYFGEISILNIKGSK 555

Qy    426 SGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLDVNAEAA 485
          :||||||||||:|||||||||:|| | |:||| |:|::|||::||:|   ||:|   |
Db    556 AGNRRTANIKSIGYSDLFCLSKDDLMEALTEYPDAKTMLEEKGKQILMKDGLLDINIANA 615

Qy    486 EIALQEATESRLRGLDQQLDDLQTKFARLLAELESSALKIAYRIERLE 533
            ::  | ::  ::  :| |||:|||:||| ||   |:  |: ::|
Db    616 GSDPKD-LEEKVTRMEGSVDLLQTRFARILAEYESMQQKLKQRLTKVE 662


RESULT  11
I50680
alpha subunit of rod photoreceptor CNG-channel - chicken
C;Species: Gallus gallus (chicken)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 13-Aug-1999
C;Accession: I50680
R;Bonigk, W.; Altenhofen, W.; Muller, F.; Dose, A.; Illing, M.; Molday, R.S.; Kaupp, U.B
Neuron 10, 865-877, 1993
A;Title: Rod and cone photoreceptor cells express distinct genes for cGMP-gated channels
A;Reference number: I50630; MUID:93264082
A;Accession: I50680
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-645 <BON>
A;Cross-references: EMBL:X89599; NID:g908852; PIDN:CAA61758.1; PID:g908853
C;Superfamily: cyclic nucleotide-gated channel; cAMP receptor protein cyclic nucleotide-
F;431-555/Domain: cAMP receptor protein cyclic nucleotide-binding domain homology <CAP>


  Query Match             51.2%;  Score 1529;  DB 2;  Length 645;
  Best Local Similarity   50.7%;  Pred. No. 1.3e-98;
  Matches  285;  Conservative  124;  Mismatches  143;  Indels   10;  Gaps    5;

Qy      6 KVKTTESSPPAPSKARKLLPVLDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHGY 65
          | |  :       : :| : ::||:|: || ||  :  |||||  ::: |||| :||: |
Db     89 KHKNKDKKKGKEEEKKKDIFIIDPAGNMYYNWLFCITMPVMYNWTMIIARACFDELQNDY 148

Qy     66 LVAWLVLDYTSDLLYLLDMVVRFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLASLMPT 125
          |  | ::|| ||::|:  || ||  ||:||||:|| :: ::  :|  :  | ||  |::||
Db    149 LAVWFIVDYVSDVIYIADMFVRTRTGYLEQGLLVKEEQKLKEKYKSSLQFKLDFLSIIPT 208

Qy    126 DVVYVRLGPHTPTLRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSC 185
          |::| :|| : | ||:|| ||  |:|| | |||||| ||| |||: |::|| ::||||:|
Db    209 DLLYFKLGLNYPELRINRLLRVARMFEFFQRTETRTNYPNIFRISNLVMYIVIIIHWNAC 268

Qy    186 LYFALSRYLGFGRDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAREEEYL 245
          :|:::|: :||| | |||| : | | || |:|:|| |:||| |||:|:|||| |: ||
Db    269 VYYSISKAIGFGADTWVYPNTSHPEFARLTRKYVYSLYWSTLTLTTIGETPPPVRDSEYF 328

Qy    246 FMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVID 305
          |:| |||: |: ||||:|:: |:| ||| | | |      :|:||  ::|:: :|:|||
Db    329 FVVVDFLVGVLIFATIVGNVGSMISNMNAARAEFQAKIDAIKQYMHFRNVSKDMEKRVIK 388

Qy    306 WYQHLQINKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQ 365
          |: :|  ||| :|  :|::||::|||||:|::||| || :|:|| :||| || |||||||
Db    389 WFDYLWTNKKAVDEREVLKYLPDKLRAEIAINVHLETLKKVRIFADCEAGLLVELVLKLQ 448

Qy    366 PQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIINIKGNM 425
          || ||||:|:|||||||:|||||:||:|||||||||: ||  | |||||||:||||:
Db    449 PQVYSPGDYICRKGDIGREMYIIKEGKLAVVADDGVTQFVVLSDGSYFGEISILNIKGSK 508

Qy    426 SGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLDVNAEAA 485
          :|||||||:|:|||||||||:|| | |:||| |: ::||||::||:|   ||:  | |
Db    509 AGNRRTANIRSIGYSDLFCLSKDDLMEALTEYPDAKAMLEEKGKQILMKDGLLDI--EVA 566

Qy    486 EIALQ-EATESRLRGLDQQLDDLQTKFARLLAELESSALKIAYRIERLEWQTREWPMPE- 543
```

```
          :    :  | ::  ::  :| |||||||||||| :::  |:  |: ::|   :  |: |
Db    567 NLGSDPKDLEEKVAYMEGSMDRLQTKFARLLAEYDAAQQKLKKRLTQIEKILK--PVMEQ 624

Qy    544 ---DLAEADDEGEPEEGTSKDE 562
            |  |||   : : | :| |
Db    625 EFLDFEEADPPTD-KPGVTKTE 645


RESULT  12
A42161
cGMP-gated cation channel, rod photoreceptor - human
C;Species: Homo sapiens (man)
C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 13-Aug-1999
C;Accession: A42161
R;Pittler, S.J.; Lee, A.K.; Altherr, M.R.; Howard, T.A.; Seldin, M.F.; Hurwitz, R.L.;
J. Biol. Chem. 267, 6257-6262, 1992
A;Title: Primary structure and chromosomal localization of human and mouse rod photor
A;Reference number: A42161; MUID:92210603
A;Accession: A42161
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-690 <PIT>
A;Cross-references: GB:M84741; NID:g180461; PIDN:AAA52010.1; PID:g180462
C;Superfamily: cyclic nucleotide-gated channel; cAMP receptor protein cyclic nucleoti
C;Keywords: cGMP binding; transmembrane protein
F;479-603/Domain: cAMP receptor protein cyclic nucleotide-binding domain homology <CA


  Query Match             51.1%;  Score 1526.5;  DB 2;  Length 690;
  Best Local Similarity   53.8%;  Pred. No. 2.1e-98;
  Matches  284;  Conservative  112;  Mismatches  131;  Indels    1;  Gaps    1;

Qy      6 KVKTTESSPPAPSKARKLLPVLDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHGY 65
          | |  | |        :| : |:|||| || ||  :  |||||  ::: |||| :||  |
Db    137 KKKKEEKSKDKKEHHKKEVVVIDPSGNTYYNWLFCITLPVMYNWTMVIARACFDELQSDY 196

Qy     66 LVAWLVLDYTSDLLYLLDMVVRFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLASLMPT 125
          |  ||:||| ||::||:|| ||  ||:||||:|| :: :: ::|     | ||: ||:||
Db    197 LEYWLILDYVSDIVYLIDMFVRTRTGYLEQGLLVKEELKLINKYKSNLQFKLDVLSLIPT 256

Qy    126 DVVYVRLGPHTPTLRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSC 185
          |::| :|| : | :|||| ||  |:|| | |||||| ||| |||: |::|| ::||||:|
Db    257 DLLYFKLGWNYPEIRLNRLLRFSRMFEFFQRTETRTNYPNIFRISNLVMYIVIIIHWNAC 316

Qy    186 LYFALSRYLGFGRDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAREEEYL 245
          :::::|: :||| | ||||    | | || |:|:|| |:||| |||:|:|||| |: ||:
Db    317 VFYSISKAIGFGNDTWVYPDINDPEFGRLARKYVYSLYWSTLTLTTIGETPPPVRDSEYV 376

Qy    246 FMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVID 305
          |:| |||: |: ||||:|:: |:| ||| | | |      :|:||  ::|:: :|:|||
Db    377 FVVVDFLIGVLIFATIVGNIGSMISNMNAARAEFQARIDAIKQYMHFRNVSKDMEKRVIK 436

Qy    306 WYQHLQINKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQ 365
          |: :|  ||| :|  :|::||::|||||:|::||| || :|:|| :||| || |||||||
Db    437 WFDYLWTNKKTVDEKEVLKYLPDKLRAEIAINVHLDTLKKVRIFADCEAGLLVELVLKLQ 496

Qy    366 PQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIINIKGNM 425
          || ||||:|:|:|||||:|||||:||:|||||||||: ||  | |||||||:||||:
Db    497 PQVYSPGDYICKKGDIGREMYIIKEGKLAVVADDGVTQFVVLSDGSTFGEISILNIKGSK 556

Qy    426 SGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLDVNAEAA 485
          :|||||||:|:|||||||||:|| | |:||| |:|::|||::||:|   ||:|   |
Db    557 AGNRRTANIKSIGYSDLFCLSKDDLMEALTEYPDAKTMLEEKGKQILMKDGLLDLNIANA 616

Qy    486 EIALQEATESRLRGLDQQLDDLQTKFARLLAELESSALKIAYRIERLE 533
            ::  | ::  ::  :| |||:|||:||| ||   |:  |: ::|
Db    617 GSDPKD-LEEKVTRMEGSVDLLQTRFARILAEYESMQQKLKQRLTKVE 663


RESULT  13
B42161
cGMP-gated cation channel, rod photoreceptor - mouse
C;Species: Mus musculus (house mouse)
```

```
C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 16-Jul-1999
C;Accession: B42161
R;Pittler, S.J.; Lee, A.K.; Altherr, M.R.; Howard, T.A.; Seldin, M.F.; Hurwitz, R.L.; Wa
J. Biol. Chem. 267, 6257-6262, 1992
A;Title: Primary structure and chromosomal localization of human and mouse rod photorece
A;Reference number: A42161; MUID:92210603
A;Accession: B42161
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-688 <PIT>
A;Cross-references: GB:M84742
A;Note: authors translated the codon TAT for residue 544 as Thr
C;Superfamily: cyclic nucleotide-gated channel; cAMP receptor protein cyclic nucleotide-
C;Keywords: cGMP binding
F;475-599/Domain: cAMP receptor protein cyclic nucleotide-binding domain homology <CAP>

  Query Match             51.0%;  Score 1525.5;  DB 2;  Length 688;
  Best Local Similarity   51.9%;  Pred. No. 2.4e-98;
  Matches  289;  Conservative  116;  Mismatches  137;  Indels   15;  Gaps    4;

Qy       6 KVKTTESSPPAPSKARKLLPVLDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHGY 65
          | |  | :     : :| : |:|||||: || ||  :  |||||  ::: |||| :||  |
Db     133 KKKKEEKTKEKKEEEKKEVVVIDPSGNTYYNWLFCITLPVMYNWTMIIARACFDELQSDY 192

Qy      66 LVAWLVLDYTSDLLYLLDMVVRFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLASLMPT 125
          |  ||: || |:::|| || ||  ||:||||:|| |: ::  :|     | ||: |::||
Db     193 LEYWLIFDYVSNVVYLADMFVRTRTGYLEQGLLVKDRMKLIEKYKANLQFKLDVLSVIPT 252

Qy     126 DVVYVRLGPHTPTLRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSC 185
          |::|:: | : | :|||| ||  |:|| | |||||| ||| |||: |:::|| ::||||:|
Db     253 DLLYIKFGWNYPEIRLNRLLRISRMFEFFQRTETRTNYPNIFRISNLVMYIVIIIHWNAC 312

Qy     186 LYFALSRYLGFGRDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAREEEYL 245
          :|:::|: :||| | ||||    | | || |:|:|| |:||| |||:|:|||  : ||:
Db     313 VYYSISKAIGFGNDTWVYPDVNDPEFGRLARKYVYSLYWSTLTLTTIGETPPPVLDSEYI 372

Qy     246 FMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVID 305
          |:| |||: |: ||||:|:: |:| ||| | | |       :|:||  ::|:: :|:|||
Db     373 FVVVDFLIGVLIFATIVGNIGSMISNMNAARAEFQSRVDAIKQYMNFRNVSKDMEKRVIK 432

Qy     306 WYQHLQINKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQ 365
          |: :|  |||  :|  :|::||::||||:|::||| || :|:|| :||| || |||||||
Db     433 WFDYLWTNKKTVDEREVLRYLPDKLRAEIAINVHLDTLKKVRIFADCEAGLLVELVLKLQ 492

Qy     366 PQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIINIKGNM 425
          || ||||:|:|:|||||:|||||:||:||||||||||: ||  | |||||||:||||:
Db     493 PQVYSPGDYICKKGDIGREMYIIKEGKLAVVADDGITQFVVLSDGSYFGEISILNIKGSK 552

Qy     426 SGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLDVN---- 481
          :|||||||||:|||||||||:|| |||:||| |:|::|||||:||:|   ||:|
Db     553 AGNRRTANIKSIGYSDLFCLSKDDLMEVLTEYPDAKTMLEEKGRQILMKDGLLDINIANM 612

Qy     482 -AEAAEIALQEATESRLRGLDQQLDDLQTKFARLLAELESSALKIAYRIERLEWQTREWP 540
           ::  ::      | ::  ::  :| |||:|| :||| ||   |:  |: ::|   :  |
Db     613 GSDPKDL------EEKVTRMEGSVDLLQTRFACILAEYESMQQKLKQRLTKVEKFLK--P 664

Qy     541 MPEDLAEADDE--GEPE 555
          : |    | :|  || |
Db     665 LIETEFSAIEEPGGESE 681


RESULT  14
JH0560
cyclic nucleotide-gated channel - channel catfish
C;Species: Ictalurus punctatus (channel catfish)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: JH0560
R;Goulding, E.H.; Ngai, J.; Kramer, R.H.; Colicos, S.; Axel, R.; Siegelbaum, S.A.; Chess
Neuron 8, 45-58, 1992
A;Title: Molecular cloning and single-channel properties of the cyclic nucleotide-gated
A;Reference number: JH0560; MUID:92110008
A;Accession: JH0560
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-682 <GOU>
A;Cross-references: GB:M83111
A;Experimental source: olfactory epithelium
C;Comment: This cyclic nucleotide-gated channel is activated equally well by both cAM
C;Superfamily: cyclic nucleotide-gated channel; cAMP receptor protein cyclic nucleoti
C;Keywords: cAMP binding; cGMP binding; ion channel; ion transport; olfaction; transm
F;137-157/Domain: transmembrane #status predicted <TS1>
F;173-193/Domain: transmembrane #status predicted <TS2>
F;217-236/Domain: transmembrane #status predicted <TS3>
F;241-261/Domain: transmembrane #status predicted <TS4>
F;277-297/Domain: transmembrane #status predicted <TS5>
F;319-337/Domain: transmembrane #status predicted <TH5>
F;350-370/Domain: transmembrane #status predicted <TS6>
F;447-571/Domain: cAMP receptor protein cyclic nucleotide-binding domain homology <CA

  Query Match             48.1%;  Score 1436.5;  DB 1;  Length 682;
  Best Local Similarity   49.8%;  Pred. No. 3.8e-92;
  Matches  288;  Conservative  100;  Mismatches  165;  Indels   25;  Gaps    6;

Qy      11 ESSPPAPSKARKLLP-----VLDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHGY 65
           :|: || :. :          |:  | | ||:||  :    :|| |:|| |||| ||
Db     110 QSAAPADAPKKTFKERWEGFVVSQSDDIYYYWLFFIALASLYNWIMLVARACFDQLQDEN 169

Qy      66 LVAWLVLDYTSDLLYLLDMVVRFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLASLMPT 125
             |: |||  |::|:||  :|  ||:||||:|| |  ::   |:||  | ||  |::||
Db     170 FFLWVGLDYLCDVIYILDTCIRLRTGYLEQGLLVKDLAKLRDNYIRTLQFKLDFLSILPT 229

Qy     126 DVVYVRLGPHTPTLRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSC 185
           ::::   | : | || || ||  |:|| |||||||| |||||||  |:||| |:||||:|
Db     230 ELLFFVTG-YVPQLRFNRLLRFSRMFEFFDRTETRTNYPNAFRICNLILYILVIIHWNAC 288

Qy     186 LYFALSRYLGFGRDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAREEEYL 245
           :|:|:|: ||   | |||        : |   |:| ||:||| |||:|: ||| ::|||:
Db     289 IYYAISKALGLSSDTWVYSGQN----KTLSFCYVYCFYWSTLTLTTIGEMPPPVKDEEYV 344

Qy     246 FMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVID 305
           |:| |||: |: ||||:|:: |:| |||   | |      :| ||  : ||| || |||
Db     345 FVVFDFLVGVLIFATIVGNVGSMIANMNATRAEFQTRIDAIKHYMHFRKVNRTLETRVIK 404

Qy     306 WYQHLQINKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQ 365
           |: :|  |||  :|  :|::||::||||:|::||| || :|:|||:||| || ||||||:
Db     405 WFDYLWTNKKTVDEQEVLKNLPDKLRAEIAINVHLDTLKKVRIFQDCEAGLLVELVLKLR 464

Qy     366 PQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIINIKGNM 425
           || ||||:|:|||||||:|||||:|||||||||||:||:|:|  ||  |||||||:||:|:
Db     465 PQVYSPGDYICRKGDIGKEMYIIKEGQLAVVADDGVTQFALLTAGGCFGEISILNIQGSK 524

Qy     426 SGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLDVNAEAA 485
           ||||||||:|:|||||||||:|| | ::||| || ::||:|||| |   || :  |
Db     525 MGNRRTANIRSIGYSDLFCLSKDDLMEAVAEYPDAQKVLEERGREILRKQGLLDESVAAG 584

Qy     486 EIALQEATESRLRGLDQQLDDLQTKFARLLAELESSALKIAYRIERLEWQ---TREWPMP 542
           : : : || ::  ||  || ||||:|||| |  |:  ::  ||  || |   |    :
Db     585 GLGVID-TEEKVERLDASLDILQTRFARLLGEFTSTQRRLKQRITALERQLCHTGLGLLS 643

Qy     543 EDLAE-----------ADDEGEPEEGTSKDEEGRASQE 569
           :: ||           ||   :|| |    |  : :|
Db     644 DNEAEGEHAGVPTHTHADIHAQPETHTRTSAETNSEEE 681


RESULT  15
S52072
DmCNGC protein - fruit fly (Drosophila sp.)
C;Species: Drosophila sp.
C;Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 16-Jul-1999
C;Accession: S52072
R;Baumann, A.; Frings, S.; Godde, M.; Seifert, R.; Kaupp, U.B.
EMBO J. 13, 5040-5050, 1994
A;Title: Primary structure and functional expression of a Drosophila cyclic nucleotid
A;Reference number: S52072; MUID:95045396
```

```
A;Accession: S52072
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-665 <BAU>
C;Genetics:
A;Gene: FlyBase:Cng
A;Cross-references: FlyBase:FBgn0014462
C;Superfamily: cyclic nucleotide-gated channel; cAMP receptor protein cyclic nucleotide-
F;429-553/Domain: cAMP receptor protein cyclic nucleotide-binding domain homology <CAP>


  Query Match            41.3%;  Score 1235;  DB 2;  Length 665;
  Best Local Similarity  47.1%;  Pred. No. 4e-78;
  Matches  243;  Conservative   98;  Mismatches  165;  Indels   10;  Gaps    2;

Qy      27 LDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHGYLVAWLVLDYTSDLLYLLDMVV 86
           :||:   :| ||  :   |:||:| :| || | ::      |  |||  | :|||| :|
Db     100 VDPTLQSHYRWLAIVSLAVLYNIIFVVGRAVFWEINKSAPAFWYTLDYLCDFIYLLDTLV 159

Qy      87 RFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLASLMPTDVVYVRLGPHTPT-------- 138
             | |||:||:|| |  |:   |  |  ::||: |::|||: |:   | | :
Db     160 HMHEGFLDQGLLVRDAFRLRRHYFHTKGWYLDVLSMLPTDLAYIWWPPETCSSLYLPCPV 219

Qy     139 -LRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSCLYFALSRYLGFG 197
            :|||| ||  ||:| |||||| | ||||||| |::| | |:||||:|:|||:|  :||
Db     220 IVRLNRLLRINRLWEWFDRTETATGYPNAFRICKVVLAILVLIHWNACMYFAISYEIGFS 279

Qy     198 RDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAREEEYLFMVGDFLLAVMG 257
             |:|||          |:|||:||||:||| |||:|:|| |  : ||||:| |||  |:
Db     280 SDSWVYNLNGTRN-NTLQRQYIYSFYWSTLTLTTIGETPTPENDVEYLFVVADFLAGVLI 338

Qy     258 FATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVIDWYQHLQINKKMT 317
           ||||:|:: |:| ||| |   |      ||:||  : |  :|| ||| |: :
Db     339 FATIVGNIGSMISNMNVARVEFQNRMDGVKQYMAFRRVGHELEARVIRWFAYTWSQSGAL 398

Qy     318 NEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQPQTYSPGEYVCR 377
           :|  :|  ||::|:||:|: ||: || :|:|| : |  ||| ||||: | :|||:|:||
Db     399 DEERVLAALPDKLKAEIAIQVHMDTLKQVRIFHDTEPGLLEALVLKLKLQVFSPGDYICR 458

Qy     378 KGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIINIKGNMSGNRRTANIKSL 437
           |||:|:|||||:: |:|:|| |||||  | ||||  |||:|:: | || :|||||||::||
Db     459 KGDVGKEMYIVKRGKLSVVGDDGITVLATLGAGSVFGEVSVLEIAGNRTGNRRTANVRSL 518

Qy     438 GYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLDVNAEAAEIALQEATESRL 497
           ||||||||:| || | ||:||:|:: : ::| ::| |   ||    |    : :: |  :
Db     519 GYSDLFCLAKRDLWETLSDYPEARSTLTQRGCQLLRKDGLLDEQIFADSQRVHDSIEGGI 578

Qy     498 RGLDQQLDDLQTKFARLLAELESSALKIAYRIERLE 533
             |:  :::|  : ||||||  :|  ||  |: :||
Db     579 EKLELSVENLNMRLARLLAEYTASQAKIKQRLAKLE 614


Search completed: June 20, 2002, 17:07:56
Job time: 8884 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 20, 2002, 17:06:57 ; Search time 37.21 Seconds
(without alignments)
598.327 Million cell updates/sec

Title: US-09-927-267-1
Perfect score: 2989
Sequence: 1 MSQDTKVKTTESSPPAPSKA.....EGTSKDEEGRASQEGPPGPE 575

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2811	94.0	575	1	CNGX_RAT	Q64359 rattus norv
2	1579	52.8	706	1	CNG3_BOVIN	Q29441 bos taurus
3	1573.5	52.6	631	1	CNG3_MOUSE	Q9jjz8 mus musculu
4	1568	52.5	663	1	CNG2_BOVIN	Q03041 bos taurus
5	1562.5	52.3	735	1	CNG1_CHICK	Q90805 gallus gall
6	1562	52.3	664	1	CNG2_RABIT	Q28718 oryctolagus
7	1558	52.1	664	1	CNG2_RAT	Q00195 rattus norv
8	1549.5	51.8	694	1	CNG3_HUMAN	Q16281 homo sapien
9	1547	51.8	664	1	CNG2_MOUSE	Q62398 mus musculu
10	1544.5	51.7	690	1	CNG1_BOVIN	Q00194 b cgmp-gate
11	1538.5	51.5	686	1	CNG1_HUMAN	P29973 h cgmp-gate
12	1533.5	51.3	684	1	CNG1_MOUSE	P29974 m cgmp-gate
13	1532.5	51.3	691	1	CNG1_CANFA	Q28279 c cgmp-gate
14	1532	51.3	683	1	CNG1_RAT	Q62927 r cgmp-gate
15	1529	51.2	645	1	CNG3_CHICK	Q90980 gallus gall
16	1436.5	48.1	682	1	CNG_ICTPU	P55934 ictalurus p
17	1235	41.3	665	1	CNG_DROME	Q24278 drosophila
18	1121.5	37.5	741	1	TAX4_CAEEL	Q03611 caenorhabdi
19	847	28.3	261	1	CNG2_HUMAN	Q16280 homo sapien
20	738	24.7	1394	1	CNG4_BOVIN	Q28181 bos taurus
21	719	24.1	909	1	CNG4_HUMAN	Q14028 homo sapien
22	377.5	12.6	1174	1	CIKE_DROME	Q02280 drosophila
23	375	12.5	1159	1	HERG_HUMAN	Q12809 homo sapien
24	159.5	5.3	768	1	KGP1_DROME	Q03042 drosophila
25	133	4.4	377	1	KAPR_APLCA	P31319 aplysia cal
26	129.5	4.3	686	1	KGPB_BOVIN	P21136 bos taurus
27	129.5	4.3	686	1	KGPB_HUMAN	P14619 homo sapien
28	129.5	4.3	686	1	KGPB_MOUSE	Q9z0z0 mus musculu
29	127.5	4.3	379	1	KAP0_BOVIN	P00514 bos taurus
30	126	4.2	327	1	KAPR_DICDI	P05987 dictyosteli
31	125.5	4.2	380	1	KAP0_RAT	P09456 rattus norv
32	125.5	4.2	670	1	KGPA_BOVIN	P00516 bos taurus
33	124.5	4.2	670	1	KGPA_HUMAN	Q13976 homo sapien
34	122.5	4.1	379	1	KAP0_PIG	P07802 sus scrofa
35	121.5	4.1	412	1	KAPR_SCHPO	P36600 schizosacch
36	121.5	4.1	670	1	KGPA_RABIT	O77676 oryctolagus
37	119.5	4.0	486	1	K1C0_XENLA	P02537 xenopus lae
38	119	4.0	390	1	KAPR_MAGGR	O14448 magnaporthe
39	118.5	4.0	762	1	KGP2_HUMAN	Q13237 homo sapien
40	117	3.9	762	1	KGP2_MOUSE	Q61410 mus musculu
41	117	3.9	762	1	KGP2_RAT	Q64595 rattus norv
42	116.5	3.9	381	1	KAP0_HUMAN	P10644 homo sapien
43	116.5	3.9	412	1	KAPR_EMENI	O59922 emericella
44	116.5	3.9	933	1	KGP3_DROME	P32023 drosophila
45	116.5	3.9	1088	1	KGP2_DROME	Q03043 drosophila

ALIGNMENTS

```
RESULT   1
CNGX_RAT
ID   CNGX_RAT       STANDARD;      PRT;   575 AA.
AC   Q64359;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Cyclic-nucleotide-gated olfactory channel OCNC2 subunit.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY;
RX   MEDLINE=94377458; PubMed=7522325;
RA   Bradley J., Li J., Davidson N., Lester H.A., Zinn K.;
RT   "Heteromeric olfactory cyclic nucleotide-gated channels: a subunit
RT   that confers increased sensitivity to cAMP.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:8890-8894(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Olfactory neuroepithelium;
RX   MEDLINE=95000663; PubMed=7522482;
RA   Liman E.R., Buck L.B.;
RT   "A second subunit of the olfactory cyclic nucleotide-gated channel
RT   confers high sensitivity to cAMP.";
RL   Neuron 13:611-621(1994).
RN   [3]
RP   SEQUENCE OF 7-35 FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY;
RA   Bradley J., Zhang Y., Bakin R., Lester H.A., Ronnett G., Zinn K.;
RL   Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: ODORANT SIGNAL TRANSDUCTION IS PROBABLY MEDIATED
CC       BY A G-PROTEIN COUPLED CASCADE USING CAMP AS SECOND MESSENGER.
CC       THE OLFACTORY CHANNEL CAN BE SHOWN TO BE ACTIVATED BY CYCLIC
CC       NUCLEOTIDES WHICH LEADS TO A DEPOLARIZATION OF OLFACTORY
CC       SENSORY NEURONS.
CC   -!- SUBUNIT: HETEROOLIGOMER OF OCNC1 AND OCNC2 SUBUNITS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: OLFACTORY NEURONS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U12623; AAA21464.1; -.
DR   EMBL; U12425; AAA64748.1; -.
DR   EMBL; U76219; AAC17596.1; -.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR000595; cNMP_binding.
DR   Pfam; PF00027; cNMP_binding; 1.
```

```
DR   Pfam; PF00520; ion_trans; 1.
DR   SMART; SM00100; cNMP; 1.
DR   PROSITE; PS00888; CNMP_BINDING_1; 1.
DR   PROSITE; PS00889; CNMP_BINDING_2; 1.
DR   PROSITE; PS50042; CNMP_BINDING_3; 1.
KW   Ionic channel; Ion transport; cAMP-binding; Transmembrane;
KW   Multigene family; Olfaction.
FT   DOMAIN        1     33       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     34     54       H1 (POTENTIAL).
FT   DOMAIN       55     65       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     66     86       H2 (POTENTIAL).
FT   DOMAIN       87    112       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    113    133       H3 (POTENTIAL).
FT   DOMAIN      134    168       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    169    189       H4 (POTENTIAL).
FT   DOMAIN      190    244       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    245    265       H5 (POTENTIAL).
FT   DOMAIN      266    400       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    401    421       H6 (POTENTIAL).
FT   DOMAIN      422    575       CYTOPLASMIC (POTENTIAL).
FT   NP_BIND     356    500       CAMP (BY SIMILARITY).
FT   BINDING     415    415       CAMP (POTENTIAL).
FT   BINDING     430    430       CAMP (POTENTIAL).
SQ   SEQUENCE   575 AA;  65674 MW;  FCF01C563214F575 CRC64;


  Query Match             94.0%;  Score 2811;  DB 1;  Length 575;
  Best Local Similarity   93.4%;  Pred. No. 4e-186;
  Matches  537;  Conservative   16;  Mismatches   22;  Indels    0;  Gaps    0;

Qy       1 MSQDTKVKTTESSPPAPSKARKLLPVLDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPD 60
           |||| |||||||:||||:|||| ||||||||||||||||||||:|||||||:||||||||
Db       1 MSQDGKVKTTESTPPAPTKARKWLPVLDPSGDYYYWWLNTMVFPIMYNLIIVVCRACFPD 60

Qy      61 LQHGYLVAWLVLDYTSDLLYLLDMVVRFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLA 120
           ||| ||||| |||||||||||||: ||||||||||||||||| |:|||||||| ||||
Db      61 LQHSYLVAWFVLDYTSDLLYLLDIGVRFHTGFLEQGILVVDKGMIASRYVRTWSFLLDLA 120

Qy     121 SLMPTDVVYVRLGPHTPTLRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVI 180
           ||:|||  ||:|||| |||||||||| ||||||||||||||||||||||||||||||||||
Db     121 SLVPTDAAYVQLGPHIPTLRLNRFLRVPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVI 180

Qy     181 HWNSCLYFALSRYLGFGRDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAR 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||| | |
Db     181 HWNSCLYFALSRYLGFGRDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPLPDR 240

Qy     241 EEEYLFMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLE 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||::||
Db     241 EEEYLFMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNKRLE 300

Qy     301 RRVIDWYQHLQINKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEEL 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 RRVIDWYQHLQINKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEEL 360

Qy     361 VLKLQPQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIIN 420
           ||||||||||||||||||||:||||||||||||||||:||||||||||||||||||||
Db     361 VLKLQPQTYSPGEYVCRKGDIGREMYIIREGQLAVVADDGVTQYAVLGAGLYFGEISIIN 420

Qy     421 IKGNMSGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLDV 480
           |||||||||||||||||||||||||||||||||||||||| :||||||||||||||||||
Db     421 IKGNMSGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQAQAVMEEKGREILLKMNKLDV 480

Qy     481 NAEAAEIALQEATESRLRGLDQQLDDLQTKFARLLAELESSALKIAYRIERLEWQTREWP 540
           |||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Db     481 NAEAAEIALQEATESRLKGLDQQLDDLQTKFARLLAELESSALKIAYRIERLEWQTREWP 540

Qy     541 MPEDLAEADDEGEPEEGTSKDEEGRASQEGPPGPE 575
           ||||: ||||| || |||||| ||:| | || | |
Db     541 MPEDMGEADDEAEPGEGTSKDGEGKAGQAGPSGIE 575


RESULT   2
CNG3_BOVIN

ID   CNG3_BOVIN     STANDARD;      PRT;   706 AA.
AC   Q29441;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Cyclic-nucleotide-gated cation channel alpha 3 (CNG channel alpha 3)
DE   (CNG-3) (CNG3) (Cyclic nucleotide gated channel alpha 3) (Cone
DE   photoreceptor cGMP-gated channel alpha subunit).
GN   CNGA3 OR CNCG3.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RX   MEDLINE=94211295; PubMed=7512693;
RA   Weyand I., Godde M., Frings S., Weiner J., Mueller F., Altenhofen W.,
RA   Hatt H., Kaupp U.B.;
RT   "Cloning and functional expression of a cyclic-nucleotide-gated
RT   channel from mammalian sperm.";
RL   Nature 368:859-863(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RX   MEDLINE=94224768; PubMed=8170936;
RA   Biel M., Zong X., Distler M., Bosse E., Klugbauer N., Murakami M.,
RA   Flockerzi V., Hofmann F.;
RT   "Another member of the cyclic nucleotide-gated channel family,
RT   expressed in testis, kidney, and heart.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:3505-3509(1994).
CC   -!- FUNCTION: COULD BE RESPONSIBLE FOR CGMP-INDUCED CALCIUM ENTRY IN
CC       CELLS OTHER THAN SENSORY CELLS. MIGHT BE INVOLVED IN CHEMOTAXIS OF
CC       SPERM.
CC   -!- SUBUNIT: FORMS FUNCTIONAL HETEROOLIGOMERIC CHANNELS WITH CNG4 IN
CC       VITRO.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: TESTIS, KIDNEY, RETINAL CONE, AND HEART.
CC   -!- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE-GATED CATION CHANNEL
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X89600; CAA61759.1; -.
DR   EMBL; X76485; CAA54023.1; -.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR000595; cNMP_binding.
DR   Pfam; PF00027; cNMP_binding; 1.
DR   Pfam; PF00520; ion_trans; 1.
DR   SMART; SM00100; cNMP; 1.
DR   PROSITE; PS00888; CNMP_BINDING_1; 1.
DR   PROSITE; PS00889; CNMP_BINDING_2; 1.
DR   PROSITE; PS50042; CNMP_BINDING_3; 1.
KW   Ionic channel; Ion transport; cAMP-binding; Transmembrane;
KW   Multigene family; Vision.
FT   TRANSMEM    186    206       POTENTIAL.
FT   TRANSMEM    322    342       POTENTIAL.
FT   TRANSMEM    398    418       POTENTIAL.
FT   NP_BIND     501    624       CAMP.
FT   BINDING     568    568       CAMP (POTENTIAL).
FT   BINDING     583    583       CAMP (POTENTIAL).
SQ   SEQUENCE   706 AA;  81132 MW;  F4990DCD29B56239 CRC64;


  Query Match             52.8%;  Score 1579;  DB 1;  Length 706;
  Best Local Similarity   52.7%;  Pred. No. 3.4e-101;
```

```
  Matches  297;  Conservative  111;  Mismatches  138;  Indels   18;  Gaps    3;

Qy       3 QDTKVKTTESSP-------PAPSKARKLLPVLDPSGDYYYWWLNTMVFPVMYNLIILVCR 55
           :| | |  |          |   : :|   |:||| : || ||  :  || ||  :|||| 
Db     149 KDDKAKKEEKEKKEEKKENPKKEEKKKDSVVMDPSSNMYYHWLTVIAVPVFYNWCLLVCR 208

Qy      56 ACFPDLQHGYLVAWLVLDYTSDLLYLLDMVVRFHTGFLEQGILVVDKGRISSRYVRTWSF 115
           ||| :||  :|: |||||||::|:|| :||:||  |||||||::|:|  |:   | :|  | 
Db     209 ACFDELQSEHLMLWLVLDYSADILYGMDMLVRARTGFLEQGLMVMDASRLWKHYTQTLHF 268

Qy     116 FLDLASLMPTDVVYVRLGPHTPTLRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLY 175
            ||: ||:|||: | :|| : | || || |:  |||| ||||||| ||| |||  |:|| 
Db     269 KLDVLSLVPTDLAYFKLGMNYPELRFNRLLKLARLFEFFDRTETRTNYPNMFRIGNLVLY 328

Qy     176 IFVVIHWNSCLYFALSRYLGFGRDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDT 235
           | ::|||||:|:|||:|:::||| |:|||| : | : || |:|:|| |:||| |||:|:| 
Db     329 ILIIIHWNACIYFAISKFIGFGTDSWVYPNVSNPEYGRLSRKYIYSLYWSTLTLTTIGET 388

Qy     236 PPPAREEEYLFMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHV 295
           ||| ::|||||:| |||: |: |||||:|:: |:| ||| : | |      :|:||: : | 
Db     389 PPPVKDEEYLFVVIDFLVGVLIFATIVGNVGSMISNMNASRAEFQAKIDSIKQYMQFRKV 448

Qy     296 NRKLERRVIDWYQHLQINKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEAS 355
           : || ||| |: :|  |||  :|  :|: ||::|:||:|::||| || :|:|||:||| 
Db     449 TKDLETRVIRWFDYLWANKKTVDEKEVLKSLPDKLKAEIAINVHLDTLRKVRIFQDCEAG 508

Qy     356 LLEELVLKLQPQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGE 415
           || ||||||:|  :|||:|:|:|||||:|||||:||:|||||:|||||: ||| | |||| 
Db     509 LLVELVLKLRPAVFSPGDYICKKGDIGREMYIIKEGKLAVVAEDGITQFVVLGDGSYFGE 568

Qy     416 ISIINIKGNMSGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKM 475
           |||:|||||: |||||||||:|:|||||||||||:|| | |:|||:|:  :|||||:||:| 
Db     569 ISILNIKGSKSGNRRTANIRSIGYSDLFCLSKDDLMEALTEYPEAKKALEEKGRQILMKD 628

Qy     476 NKLD-----VNAEAAEIALQEATESRLRGLDQQLDDLQTKFARLLAELESSALKIAYRIE 530
           | :|       |:  :|      | ::  |:  || |||:|||||||  :: :|:  |: 
Db     629 NLIDEELAKAGADPKDI------EEKVEHLETSLDSLQTRFARLLAEYNATQMKVKQRLS 682

Qy     531 RLEWQTREWPMPEDLAEADDEGEP 554
           :|| | :    |:  |   :  :| 
Db     683 QLESQVKMGLPPDGDAPQTEASQP 706


RESULT   3
CNG3_MOUSE
ID   CNG3_MOUSE     STANDARD;      PRT;   631 AA.
AC   Q9JJZ8; Q9WV01;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Cyclic-nucleotide-gated cation channel alpha 3 (CNG channel alpha 3)
DE   (CNG-3) (CNG3) (Cyclic nucleotide gated channel alpha 3) (Cone
DE   photoreceptor cGMP-gated channel alpha subunit).
GN   CNGA3 OR CNG3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL; TISSUE=Retina;
RX   MEDLINE=20273944; PubMed=10813773;
RA   Hirano A.A., Hack I., Waessle H., Duvoisin R.M.;
RT   "Cloning and immunocytochemical localization of a cyclic nucleotide-
RT   gated channel alpha-subunit to all cone photoreceptors in the mouse
RT   retina.";
RL   J. Comp. Neurol. 421:80-94(2000).
RN   [2]
RP   SEQUENCE OF 95-631 FROM N.A.
RC   STRAIN=129/SV;
RX   MEDLINE=99307448; PubMed=10377453;
RA   Biel M., Seeliger M., Pfeifer A., Kohler K., Gerstner A., Ludwig A.,
RA   Jaissle G., Fauser S., Zrenner E., Hofmann F.;
RT   "Selective loss of cone function in mice lacking the cyclic
RT   nucleotide-gated channel CNG3.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:7553-7557(1999).
CC   -!- FUNCTION: VISUAL SIGNAL TRANSDUCTION IS MEDIATED BY A G-PROTEIN
CC       COUPLED CASCADE USING CGMP AS SECOND MESSENGER. THIS PROTEIN CAN
CC       BE ACTIVATED BY CYCLIC GMP WHICH LEADS TO A OPENING OF THE CATION
CC       CHANNEL AND THEREBY CAUSING A DEPOLARIZATION OF CONE
CC       PHOTORECEPTORS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: PROMINENTLY EXPRESSED IN RETINA.
CC   -!- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE-GATED CATION CHANNEL
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ243933; CAB89685.1; -.
DR   EMBL; AJ238239; CAB42891.1; -.
DR   EMBL; AJ238240; CAB42891.1; JOINED.
DR   EMBL; AJ238241; CAB42891.1; JOINED.
DR   MGD; MGI:1341818; Cnga3.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR001622; Channel_pore_K.
DR   InterPro; IPR000595; cNMP_binding.
DR   Pfam; PF00027; cNMP_binding; 1.
DR   Pfam; PF00520; ion_trans; 1.
DR   SMART; SM00100; cNMP; 1.
DR   PROSITE; PS00888; CNMP_BINDING_1; 1.
DR   PROSITE; PS00889; CNMP_BINDING_2; 1.
DR   PROSITE; PS50042; CNMP_BINDING_3; 1.
KW   Ionic channel; Ion transport; cAMP-binding; Transmembrane;
KW   Multigene family; Vision.
FT   TRANSMEM    112    133       POTENTIAL.
FT   TRANSMEM    244    264       POTENTIAL.
FT   TRANSMEM    320    340       POTENTIAL.
FT   NP_BIND     423    546       CAMP.
FT   BINDING     490    490       CAMP (POTENTIAL).
FT   BINDING     505    505       CAMP (POTENTIAL).
FT   CONFLICT    110    110       C -> Y (IN REF. 2).
FT   CONFLICT    157    157       L -> V (IN REF. 2).
SQ   SEQUENCE   631 AA;  72641 MW;. 4FA8CD3B9AA3FE6C CRC64;


  Query Match             52.6%;  Score 1573.5;  DB 1;  Length 631;
  Best Local Similarity   54.4%;  Pred. No. 6.9e-101;
  Matches  296;  Conservative  108;  Mismatches  129;  Indels   11;  Gaps    3;

Qy       2 SQDTKVKTTESS---------PPAPSKARKLLP-VLDPSGDYYYWWLNTMVFPVMYNLII 51
           |:  :| | ||:          ||   : ||  | |:||| : |  ||  :  || ||  : 
Db      67 SELKEVSTRESNAQPNPGEQKPPDGGEGRKEEPIVVDPSSNIYCRWLTAIALPVFYNWCL 126

Qy      52 LVCRACFPDLQHGYLVAWLVLDYTSDLLYLLDMVVRFHTGFLEQGILVVDKGRISSRYVR 111
           ||||||| :||  :|  ||||||::|:||:|||:||  |||||||::| |  |:   | : 
Db     127 LVCRACFDELQSEHLTLWLVLDYSADVLYVLDMLVRARTGFLEQGLMVRDTKRLWKHYTK 186

Qy     112 TWSFFLDLASLMPTDVVYVRLGPHTPTLRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAK 171
           |  | ||: ||:|||: |::|| : | || || |:  |||| |||||||| ||| |||
Db     187 TLHFKLDILSLIPTDLAYLKLGVNYPELRFNRLLKFSRLFEFFDRTETRTNYPNVFRIGN 246

Qy     172 LMLYIFVVIHWNSCLYFALSRYLGFGRDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTT 231
           |:||  ::|||||:|:|||:|:::||| |:|||: ::| : || |:|:|| |:||| |||
Db     247 LVLYTLIIIHWNACIYFAISKFIGFGTDSWVYPNTSKPEYARLSRKYIYSLYWSTLTLTT 306

Qy     232 VGDTPPPAREEEYLFMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMK 291
           :|:|||| ::|||||:| |||: :: ||||:|:: |:| |||      |       ||:||: 
Db     307 IGETPPPVKDEEYLFVVIDFLVGILIFATIVGNVGSMISNMNAPRVEFQAKIDSVKQYMQ 366

Qy     292 LQHVNRKLERRVIDWYQHLQINKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQN 351
```

```
          : | : || ||| |: :|  |:|  :|  :|::||::|:||:|::||| || :|:|||:
Db    367 FRKVTKDLETRVIRWFDYLWANRKTVDEKEVLKNLPDKLKAEIAINVHLDTLKKVRIFQD 426

Qy    352 CEASLLEELVLKLQPQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGL 411
          ||| || |||||:|  :|||:|:|:|||||:|||||:||:|||||||:||: ||  |
Db    427 CEAGLLVELVLKLRPTVFSPGDYICKKGDIGREMYIIKEGKLAVVADDGVTQFVVLSDGS 486

Qy    412 YFGEISIINIKGNMSGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREI 471
          |||||||:||||: |||||||||:|:|||||||||||:|| | |:||| |:  :|||||:|
Db    487 YFGEISILNIKGSKSGNRRTANIRSIGYSDLFCLSKDDLMEALTEYPDAKRALEEKGRQI 546

Qy    472 LLKMNKLDVNAEAAEIALQEATESRLRGLDQQLDDLQTKFARLLAELESSALKIAYRIER 531
          |:| | :| :  || :  ::  | ::  |:  || |||:|||||||  :| :|:  |: |
Db    547 LMKDNLIDEDLVAARVDTRD-VEEKVEYLESSLDILQTRFARLLAEYSASQMKLKQRLTR 605

Qy    532 LEWQ 535
          || |
Db    606 LESQ 609


RESULT   4
CNG2_BOVIN
ID   CNG2_BOVIN     STANDARD;      PRT;   663 AA.
AC   Q03041;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Cyclic-nucleotide-gated olfactory channel (Cyclic-nucleotide-gated
DE   cation channel 2) (CNG channel 2) (CNG-2) (CNG2).
GN   CNGA2 OR CNCG2.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Olfactory epithelium;
RX   MEDLINE=91032022; PubMed=1699793;
RA   Ludwig J., Margalit T., Eismann E., Lancet D., Kaupp U.B.;
RT   "Primary structure of cAMP-gated channel from bovine olfactory
RT   epithelium.";
RL   FEBS Lett. 270:24-29(1990).
CC   -!- FUNCTION: ODORANT SIGNAL TRANSDUCTION IS PROBABLY MEDIATED
CC       BY A G-PROTEIN COUPLED CASCADE USING CAMP AS SECOND MESSENGER.
CC       THE OLFACTORY CHANNEL CAN BE SHOWN TO BE ACTIVATED BY CYCLIC
CC       NUCLEOTIDES WHICH LEADS TO A DEPOLARIZATION OF OLFACTORY
CC       SENSORY NEURONS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: OLFACTORY NEURONS.
CC   -!- MISCELLANEOUS: THE OLFACTORY CHANNEL IS ACTIVATED BY BOTH CAMP AND
CC       CGMP AT SIMILAR CONCENTRATIONS, WHEREAS THE CGMP-GATED CHANNEL IS
CC       MUCH LESS SENSITIVE TO CAMP.
CC   -!- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE-GATED CATION CHANNEL
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X55010; CAA38754.1; -.
DR   PIR; S11521; S11521.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR000595; cNMP_binding.
DR   Pfam; PF00027; cNMP_binding; 1.
DR   Pfam; PF00520; ion_trans; 1.
DR   SMART; SM00100; cNMP; 1.
DR   PROSITE; PS00888; CNMP_BINDING_1; 1.
DR   PROSITE; PS00889; CNMP_BINDING_2; 1.
DR   PROSITE; PS50042; CNMP_BINDING_3; 1.
KW   Ionic channel; Ion transport; cAMP-binding; Transmembrane;
KW   Multigene family; Olfaction; Glycoprotein.
FT   DOMAIN        1    140       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    141    160       H1 (POTENTIAL).
FT   DOMAIN      161    173       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    174    192       H2 (POTENTIAL).
FT   DOMAIN      193    216       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    217    236       H3 (POTENTIAL).
FT   DOMAIN      237    274       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    275    297       H4 (POTENTIAL).
FT   DOMAIN      298    349       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    350    369       H5 (POTENTIAL).
FT   DOMAIN      370    453       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    454    474       H6 (POTENTIAL).
FT   DOMAIN      475    663       CYTOPLASMIC (POTENTIAL).
FT   NP_BIND     462    584       CAMP (BY SIMILARITY).
FT   BINDING     521    521       CAMP (BY SIMILARITY).
FT   BINDING     536    536       CAMP (BY SIMILARITY).
FT   CARBOHYD    379    379       N-LINKED (GLCNAC. . .) (PROBABLE).
SQ   SEQUENCE   663 AA;  76014 MW;  AB66D4F9203844EF CRC64;


  Query Match           52.5%;  Score 1568;  DB 1;  Length 663;
  Best Local Similarity 54.2%;  Pred. No. 1.8e-100;
  Matches  292;  Conservative  108;  Mismatches  133;  Indels    6;  Gaps    2;

Qy     18 SKARKLLPVLDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHGYLVAWLVLDYTSD 77
          :| :  | ||||:||:|| ||  :  ||:||  :|| |||| ||| || : |||||| ||
Db    124 TKKKFELFVLDPAGDWYYRWLFLIALPVLYNWCLLVARACFSDLQKGYYIVWLVLDYVSD 183

Qy     78 LLYLLDMVVRFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLASLMPTDVVYVRLGPHTP 137
          ::|: |: :|  |||||||:|| |  ::   |: |  | ||:||::|||::|  :| | |
Db    184 VVYIADLFIRLRTGFLEQGLLVKDTKKLRDNYIHTMQFKLDVASIIPTDLIYFAVGIHNP 243

Qy    138 TLRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSCLYFALSRYLGFG 197
           :| || |   |:|| |||||||:||| ||| |:||| ::|||:|:|:|:|: :|||
Db    244 EVRFNRLLHFARMFEFFDRTETRTSYPNIFRISNLILYILIIIHWNACIYYAISKSIGFG 303

Qy    198 RDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAREEEYLFMVGDFLLAVMG 257
           | ||||:  | :  | |:|:|  |:||| |||:|:|||| ::||||||:: |||: |:
Db    304 VDTWVYPNITDPEYGYLSREYIYCLYWSTLTLTTIGETPPPVKDEEYLFVIFDFLIGVLI 363

Qy    258 FATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVIDWYQHLQINKKMT 317
          ||||:|:: |:| |||   | |      || ||: : |::::| :|| |: :|  |||
Db    364 FATIVGNVGSMISNMNATRAEFQAKIDAVKHYMQFRKVSKEMEAKVIRWFDYLWTNKKSV 423

Qy    318 NEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQPQTYSPGEYVCR 377
           :|  :|::|| :||||:|::|||||| :|:|||:||| || |||||:|| :|||:|:||
Db    424 DEREVLKNLPAKLRAEIAINVHLSTLKKVRIFQDCEAGLLVELVLKLRPQVFSPGDYICR 483

Qy    378 KGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIINIKGNMSGNRRTANIKSL 437
          |||||:|||||:||:|||||||:||||:| ||  ||||||:||||:  |||||||:||
Db    484 KGDIGKEMYIIKEGKLAVVADDGVTQYALLSAGSCFGEISILNIKGSKMGNRRTANIRSL 543

Qy    438 GYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLDVNAEAA--EIALQEATES 495
          ||||||||||:|| | ::||| |: ::||:|||||:|   || |  ||  |: :||
Db    544 GYSDLFCLSKDDLMEAVTEYPDAKRVLEERGREILMKEGLLDENEVAASMEVDVQE---- 599

Qy    496 RLRGLDQQLDDLQTKFARLLAELESSALKIAYRIERLEWQTREWPMPEDLAEADDEGEP 554
           :|  |:  :| | |:|||||||   :  |:  ||  || : ::    : |::  :  ||
Db    600 KLEQLETNMDTLYTRFARLLAEYTGAQQKLKQRITVLETKMKQNNEDDSLSDGMNSPEP 658


RESULT   5
CNG1_CHICK
ID   CNG1_CHICK     STANDARD;      PRT;   735 AA.
AC   Q90805;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Cyclic nucleotide gated channel, cone photoreceptor, alpha subunit
DE   (CNG channel 1) (CNG-1).
```

```
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93264082; PubMed=7684234;
RA   Boenigk W., Altenhofen W., Mueller F., Dose A., Illing M.,
RA   Molday R.S., Kaupp U.B.;
RT   "Rod and cone photoreceptor cells express distinct genes for
RT   cGMP-gated channels.";
RL   Neuron 10:865-877(1993).
CC   -!- FUNCTION: VISUAL SIGNAL TRANSDUCTION IS MEDIATED BY A G-PROTEIN
CC       COUPLED CASCADE USING CGMP AS SECOND MESSENGER. THIS PROTEIN CAN
CC       BE ACTIVATED BY CYCLIC GMP WHICH LEADS TO A OPENING OF THE CATION
CC       CHANNEL AND THEREBY CAUSING A DEPOLARIZATION OF CONE
CC       PHOTORECEPTORS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE-GATED CATION CHANNEL
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X89598; CAA61757.1; -.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR000595; cNMP_binding.
DR   Pfam; PF00027; cNMP_binding; 1.
DR   Pfam; PF00520; ion_trans; 1.
DR   SMART; SM00100; cNMP; 1.
DR   PROSITE; PS00888; CNMP_BINDING_1; 1.
DR   PROSITE; PS00889; CNMP_BINDING_2; 1.
DR   PROSITE; PS50042; CNMP_BINDING_3; 1.
KW   Ionic channel; Ion transport; cAMP-binding; Transmembrane; Vision;
KW   Multigene family.
FT   DOMAIN        1    210       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    211    230       H1 (POTENTIAL).
FT   DOMAIN      231    243       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    244    262       H2 (POTENTIAL).
FT   DOMAIN      263    286       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    287    306       H3 (POTENTIAL).
FT   DOMAIN      307    344       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    345    367       H4 (POTENTIAL).
FT   DOMAIN      368    419       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    420    439       H5 (POTENTIAL).
FT   DOMAIN      440    523       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    524    544       H6 (POTENTIAL).
FT   DOMAIN      545    735       CYTOPLASMIC (POTENTIAL).
FT   NP_BIND     532    654       CAMP (BY SIMILARITY).
FT   BINDING     591    591       CAMP (POTENTIAL).
FT   BINDING     606    606       CAMP (POTENTIAL).
FT   CARBOHYD    449    449       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   735 AA;  85031 MW;  A67ADFDD942CEFCE CRC64;


  Query Match             52.3%;  Score 1562.5;  DB 1;  Length 735;
  Best Local Similarity   53.5%;  Pred. No. 4.9e-100;
  Matches  299;  Conservative  108;  Mismatches  149;  Indels    3;  Gaps    3;

Qy       3 QDTKVKTTESSPPAPSKARKLLPVLDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQ 62
Db     180 KDDK-KDDKKDDKKKEEQKKEVFVIDPSSNMYYNWLTIIAAPVFYNWCMLICRACFDELQ 238

Qy      63 HGYLVAWLVLDYTSDLLYLLDMVVRFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLASL 122
Db     239 IDHIKLWLFLDYCSDIIYVFDMFVRFRTGFLEQGLLVKDEKKLRDHYTQTVQFKLDVLSL 298

Qy     123 MPTDVVYVRLGPHTPTLRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHW 182
Db     299 LPTDLAYLKLGLNYPELRFNRLLRIARLFEFFDRTETRTNYPNMFRIGNLVLYILIIIHW 358

Qy     183 NSCLYFALSRYLGFGRDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAREE 242
Db     359 NACIYFAISKVIGFGTDSWVYPNVSIPEYGRLSRKYIYSLYWSTLTLTTIGETPPPVKDE 418

Qy     243 EYLFMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERR 302
Db     419 EYLFVVIDFLVGVLIFATIVGNVGSMISNMNASRAEFQAKVDSIKQYMHFRKVTKDLEAR 478

Qy     303 VIDWYQHLQINKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVL 362
Db     479 VIKWFDYLWTNKKTVDEKEVLKNLPDKLKAEIAINVHLDTLKKVRIFQDCEAGLLIELVL 538

Qy     363 KLQPQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIINIK 422
Db     539 KLKPTVFSPGDYICKKGDIGREMYIIKEGKLAVVADDGITQFVVLSDGSYFGEISILNIK 598

Qy     423 GNMSGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLDVNA 482
Db     599 GSKSGNRRTANIRSIGYSDLFCLSKDDLMEALTEYPEAKKALEEKGRQILMKDNLIDEEA 658

Qy     483 EAAEIALQEATESRLRGLDQQLDDLQTKFARLLAELESSALKIAYRIERLEWQTREWPMP 542
Db     659 AKAG-ADPKDLEEKIDRLETALDTLQTRFARLLAEYSSSQQKVKQRLARVETRVKKYG-S 716

Qy     543 EDLAEADDEGEPEEGTSKD 561
Db     717 GSLSVGEPEPEKPEEQKKD 735


RESULT   6
CNG2_RABIT
ID   CNG2_RABIT     STANDARD;      PRT;   664 AA.
AC   Q28718;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Cyclic-nucleotide-gated olfactory channel (Cyclic-nucleotide-gated
DE   cation channel 2) (CNG channel 2) (CNG-2) (CNG2) (Aorta CNG channel)
DE   (RACNG).
GN   CNGA2 OR CNCG2.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Aorta;
RX   MEDLINE=93359035; PubMed=7689061;
RA   Biel M., Altenhofen W., Hullin R., Ludwig J., Freichel M.,
RA   Flockerzi V., Dascal N., Kaupp U.B., Hofmann F.;
RT   "Primary structure and functional expression of a cyclic nucleotide-
RT   gated channel from rabbit aorta.";
RL   FEBS Lett. 329:134-138(1993).
CC   -!- FUNCTION: ODORANT SIGNAL TRANSDUCTION IS PROBABLY MEDIATED
CC       BY A G-PROTEIN COUPLED CASCADE USING CAMP AS SECOND MESSENGER.
CC       THE OLFACTORY CHANNEL CAN BE SHOWN TO BE ACTIVATED BY CYCLIC
CC       NUCLEOTIDES WHICH LEADS TO A DEPOLARIZATION OF OLFACTORY
CC       SENSORY NEURONS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE-GATED CATION CHANNEL
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
CC   ----------------------------------------------------------------------
DR   EMBL; X59668; CAA42201.1; ALT_INIT.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR000595; cNMP_binding.
DR   Pfam; PF00027; cNMP_binding; 1.
DR   Pfam; PF00520; ion_trans; 1.
DR   SMART; SM00100; cNMP; 1.
DR   PROSITE; PS00888; CNMP_BINDING_1; 1.
DR   PROSITE; PS00889; CNMP_BINDING_2; 1.
DR   PROSITE; PS50042; CNMP_BINDING_3; 1.
KW   Ionic channel; Ion transport; cAMP-binding; Transmembrane;
KW   Multigene family; Olfaction.
FT   DOMAIN        1    140       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    141    160       H1 (POTENTIAL).
FT   DOMAIN      161    173       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    174    192       H2 (POTENTIAL).
FT   DOMAIN      193    216       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    217    236       H3 (POTENTIAL).
FT   DOMAIN      237    274       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    275    297       H4 (POTENTIAL).
FT   DOMAIN      298    349       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    350    369       H5 (POTENTIAL).
FT   DOMAIN      370    453       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    454    474       H6 (POTENTIAL).
FT   DOMAIN      475    664       CYTOPLASMIC (POTENTIAL).
FT   NP_BIND     462    584       CAMP (BY SIMILARITY).
FT   BINDING     521    521       CAMP (POTENTIAL).
FT   BINDING     536    536       CAMP (POTENTIAL).
FT   CARBOHYD    379    379       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   664 AA;  76205 MW;  5E9170D0B322B3E9 CRC64;
```

```
  Query Match           52.3%;  Score 1562;  DB 1;  Length 664;
  Best Local Similarity 54.2%;  Pred. No. 4.6e-100;
  Matches  292;  Conservative  108;  Mismatches  133;  Indels    6;  Gaps    2;

Qy      18 SKARKLLPVLDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHGYLVAWLVLDYTSD 77
           :| :  | ||||:||:|| ||  :  ||:||  :|| |||| ||| || : |||||| ||
Db     124 TKKKFELFVLDPAGDWYYRWLFVIAMPVLYNWCLLVARACFSDLQRGYFLVWLVLDYFSD 183

Qy      78 LLYLLDMVVRFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLASLMPTDVVYVRLGPHTP 137
           ::|: |: :|  |||||||:|| |  ::   |: |  | ||:||::|||::|  :| | |
Db     184 VVYIADLFIRLRTGFLEQGLLVKDPKKLRDNYIHTLQFKLDVASIIPTDLIYFAVGIHNP 243

Qy     138 TLRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSCLYFALSRYLGFG 197
           || || |   |:|| ||||||||:||| |||: |:||| |:|||:|:|:|:|: :|||
Db     244 ELRFNRLLHFARMFEFFDRTETRTSYPNIFRISNLVLYILVIIHWNACIYYAISKSIGFG 303

Qy     198 RDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAREEEYLFMVGDFLLAVMG 257
           | ||||:   | :  | |:|:|  |:||| |||:|:|||| ::|||||:: |||: |:
Db     304 VDTWVYPNITDPEYGYLAREYIYCLYWSTLTLTTIGETPPPVKDEEYLFVIFDFLIGVLI 363

Qy     258 FATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVIDWYQHLQINKKMT 317
           ||||:|:: |:| |||   | |      || ||: : |:::: :|| |: :|  |||
Db     364 FATIVGNVGSMISNMNATRAEFQAKIDAVKHYMQFRKVSKEMEAKVIKWFDYLWTNKKTV 423

Qy     318 NEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQPQTYSPGEYVCR 377
           :|  :|:|| :||||:|::|||||| :|:|||:||| || |||||:|| :|||:|:||
Db     424 DEREVLKNLPAKLRAEIAINVHLSTLKKVRIFQDCEAGLLVELVLKLRPQVFSPGDYICR 483

Qy     378 KGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIINIKGNMSGNRRTANIKSL 437
           |||||:|||||:||:|||||||:||||:| ||  |||||:|||||:  |||||||:||
Db     484 KGDIGKEMYIIKEGKLAVVADDGVTQYALLSAGSCFGEISILNIKGSKMGNRRTANIRSL 543

Qy     438 GYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLDVNAEAA--EIALQEATES 495
           ||||||||||:|| | ::||| |: ::||:|||||:|   || |  ||  |: :||
Db     544 GYSDLFCLSKDDLMEAVTEYPDAKKVLEERGREILMKEGLLDENEVAASMEVDVQE---- 599

Qy     496 RLRGLDQQLDDLQTKFARLLAELESSALKIAYRIERLEWQTREWPMPEDLAEADDEGEP 554
           :|: |:  :: | |:| |||||   :  |:  ||  || : ::    : |::  :  ||
Db     600 KLKQLETNMETLYTRFGRLLAEYTGAQQKLKQRITVLEVKMKQNTEDDYLSDGMNSPEP 658
```

```
RESULT   7
CNG2_RAT
ID   CNG2_RAT       STANDARD;      PRT;   664 AA.
AC   Q00195;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Cyclic-nucleotide-gated olfactory channel (Cyclic-nucleotide-gated
DE   cation channel 2) (CNG channel 2) (CNG2) (CNG-2) (OCNC1).
GN   CNGA2 OR CNCG2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Olfactory sensory neuron;
RX   MEDLINE=90370115; PubMed=1697649;
RA   Dhallan R.S., Yau K.-W., Schrader K.A., Reed R.R.;
RT   "Primary structure and functional expression of a cyclic nucleotide-
RT   activated channel from olfactory neurons.";
RL   Nature 347:184-187(1990).
CC   -!- FUNCTION: ODORANT SIGNAL TRANSDUCTION IS PROBABLY MEDIATED
CC       BY A G-PROTEIN COUPLED CASCADE USING CAMP AS SECOND MESSENGER.
CC       THE OLFACTORY CHANNEL CAN BE SHOWN TO BE ACTIVATED BY CYCLIC
CC       NUCLEOTIDES WHICH LEADS TO A DEPOLARIZATION OF OLFACTORY
CC       SENSORY NEURONS.
CC   -!- SUBUNIT: HETEROOLIGOMER OF OCNC1 AND OCNC2 SUBUNITS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: OLFACTORY NEURONS.
CC   -!- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE-GATED CATION CHANNEL
CC       FAMILY.
CC   -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-2 IS THE INITIATOR.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X55519; CAA39135.1; -.
DR   PIR; S11517; S11517.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR000595; cNMP_binding.
DR   Pfam; PF00027; cNMP_binding; 1.
DR   Pfam; PF00520; ion_trans; 1.
DR   SMART; SM00100; cNMP; 1.
DR   PROSITE; PS00888; CNMP_BINDING_1; 1.
DR   PROSITE; PS00889; CNMP_BINDING_2; 1.
DR   PROSITE; PS50042; CNMP_BINDING_3; 1.
KW   Ionic channel; Ion transport; cAMP-binding; Transmembrane;
KW   Multigene family; Olfaction.
FT   DOMAIN        1    142       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    143    162       H1 (POTENTIAL).
FT   DOMAIN      163    175       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    176    194       H2 (POTENTIAL).
FT   DOMAIN      195    218       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    219    238       H3 (POTENTIAL).
FT   DOMAIN      239    276       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    277    299       H4 (POTENTIAL).
FT   DOMAIN      300    351       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    352    371       H5 (POTENTIAL).
FT   DOMAIN      372    455       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    456    476       H6 (POTENTIAL).
FT   DOMAIN      477    664       CYTOPLASMIC (POTENTIAL).
FT   NP_BIND     464    586       CAMP (BY SIMILARITY).
FT   BINDING     523    523       CAMP (POTENTIAL).
FT   BINDING     538    538       CAMP (POTENTIAL).
FT   CARBOHYD    381    381       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT     212    212       K -> R.
SQ   SEQUENCE   664 AA;  76176 MW;  720806950EC27F3C CRC64;
```

```
  Query Match           52.1%;  Score 1558;  DB 1;  Length 664;
  Best Local Similarity  54.2%;  Pred. No. 8.7e-100;
  Matches  292;  Conservative  106;  Mismatches  135;  Indels    6;  Gaps    2;

Qy      18 SKARKLLPVLDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHGYLVAWLVLDYTSD 77
           :| :  | ||||:||:|| ||  :  ||:||  :|| |||| |||  | | |||||| ||
Db     126 TKKKFELFVLDPAGDWYYRWLFVIAMPVLYNWCLLVARACFSDLQRNYFVVWLVLDYFSD 185

Qy      78 LLYLLDMVVRFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLASLMPTDVVYVRLGPHTP 137
            :|: |:::|  |||||||:|| |  ::   |: |  | ||:||::|||::|  :| |:|
Db     186 TVYIADLIIRLRTGFLEQGLLVKDPKKLRDNYIHTLQFKLDVASIIPTDLIYFAVGIHSP 245

Qy     138 TLRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSCLYFALSRYLGFG 197
           :| || |  |:|| |||||||:||| |||: |:||| |:||||:|:|: :|: :|||
Db     246 EVRFNRLLHFARMFEFFDRTETRTSYPNIFRISNLVLYILVIIHWNACIYYVISKSIGFG 305

Qy     198 RDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAREEEYLFMVGDFLLAVMG 257
           | ||||:   | :  | |:|:|  |:||| |||:|:|||| ::|||||:: |||: |:
Db     306 VDTWVYPNITDPEYGYLAREYIYCLYWSTLTLTTIGETPPPVKDEEYLFVIFDFLIGVLI 365

Qy     258 FATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVIDWYQHLQINKKMT 317
           ||||:|:: |:| |||   | |      || ||: : |:: :| :|| |: :|  |||
Db     366 FATIVGNVGSMISNMNATRAEFQAKIDAVKHYMQFRKVSKDMEAKVIKWFDYLWTNKKTV 425

Qy     318 NEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQPQTYSPGEYVCR 377
           :|  :|:::|| :||||:|::|||||| :|:|||:||| || |||||:|| :|||:|:||
Db     426 DEREVLKNLPAKLRAEIAINVHLSTLKKVRIFQDCEAGLLVELVLKLRPQVFSPGDYICR 485

Qy     378 KGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIINIKGNMSGNRRTANIKSL 437
           |||||:|||||:||:|||||||||:||||:| ||  |||||:||||:  ||||||||:||
Db     486 KGDIGKEMYIIKEGKLAVVADDGVTQYALLSAGSCFGEISILNIKGSKMGNRRTANIRSL 545

Qy     438 GYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLDVNAEAA--EIALQEATES 495
           ||||||||||:|| | ::||| |: ::||:|||||:|   || |  ||  |: :||
Db     546 GYSDLFCLSKDDLMEAVTEYPDAKKVLEERGREILMKEGLLDENEVAASMEVDVQE---- 601

Qy     496 RLRGLDQQLDDLQTKFARLLAELESSALKIAYRIERLEWQTREWPMPEDLAEADDEGEP 554
           :|  |:  :| | |:|||||||  :  |:  ||  || : ::    : |::  :  ||
Db     602 KLEQLETNMDTLYTRFARLLAEYTGAQQKLKQRITVLETKMKQNHEDDYLSDGINTPEP 660


RESULT   8
CNG3_HUMAN
ID   CNG3_HUMAN     STANDARD;      PRT;   694 AA.
AC   Q16281; Q9UP64;
DT   01-NOV-1997 (Rel. 35, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Cyclic-nucleotide-gated cation channel alpha 3 (CNG channel alpha 3)
DE   (CNG-3) (CNG3) (Cyclic nucleotide gated channel alpha 3) (Cone
DE   photoreceptor cGMP-gated channel alpha subunit).
GN   CNGA3 OR CNCG3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98176633; PubMed=9517456;
RA   Wissinger B., Muller F., Weyand I., Schuffenhauer S., Thanos S.,
RA   Kaupp U.B., Zrenner E.;
RT   "Cloning, chromosomal localization and functional expression of the
RT   gene encoding the alpha-subunit of the cGMP-gated channel in human
RT   cone photoreceptors.";
RL   Eur. J. Neurosci. 9:2512-2521(1997).
RN   [2]
RP   SEQUENCE OF 320-580 FROM N.A.
RX   MEDLINE=95175019; PubMed=7532814;
RA   Distler M., Biel M., Flockerzi V., Hofmann F.;
RT   "Expression of cyclic nucleotide-gated cation channels in non-sensory
RT   tissues and cells.";
RL   Neuropharmacology 33:1275-1282(1994).
RN   [3]
RP   VARIANTS RMCH, AND VARIANT M-153.
RX   MEDLINE=98324775; PubMed=9662398;
RA   Kohl S., Marx T., Giddings I., Jaegle H., Jacobson S.G.,
RA   Apfelstedt-Sylla E., Zrenner E., Sharpe L.T., Wissinger B.;
RT   "Total colourblindness is caused by mutations in the gene encoding the
RT   alpha-subunit of the cone photoreceptor cGMP-gated cation channel.";
RL   Nat. Genet. 19:257-259(1998).
CC   -!- FUNCTION: VISUAL SIGNAL TRANSDUCTION IS MEDIATED BY A G-PROTEIN
CC       COUPLED CASCADE USING CGMP AS SECOND MESSENGER. THIS PROTEIN CAN
CC       BE ACTIVATED BY CYCLIC GMP WHICH LEADS TO A OPENING OF THE CATION
CC       CHANNEL AND THEREBY CAUSING A DEPOLARIZATION OF CONE
CC       PHOTORECEPTORS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: PROMINENTLY EXPRESSED IN RETINA.
CC   -!- DISEASE: DEFECTS IN CNGA3 ARE A CAUSE OF ROD MONOCHROMACY (RMCH OR
CC       RM); ALSO KNOWN AS TOTAL COLORBLINDNESS OR ACHROMATOPSIA. RMCH IS
CC       AN AUTOSOMAL RECESSIVELY INHERITED CONDITION.
CC   -!- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE-GATED CATION CHANNEL
CC       FAMILY.
CC   -!- DATABASE: NAME=Mutations of the CNGA3 gene;
CC       NOTE=Retina International's Scientific Newsletter;
CC       WWW="http://www.retina-international.com/sci-news/cnga3mut.htm".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF065314; AAC17440.1; -.
DR   EMBL; S76069; AAD14208.1; -.
DR   HSSP; P00514; 1BPK.
DR   MIM; 600053; -.
DR   MIM; 216900; -.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR001622; Channel_pore_K.
DR   InterPro; IPR000595; cNMP_binding.
DR   Pfam; PF00027; cNMP_binding; 2.
DR   Pfam; PF00520; ion_trans; 1.
DR   SMART; SM00100; cNMP; 1.
DR   PROSITE; PS00888; CNMP_BINDING_1; 1.
DR   PROSITE; PS00889; CNMP_BINDING_2; 1.
DR   PROSITE; PS50042; CNMP_BINDING_3; 1.
KW   Ionic channel; Ion transport; cAMP-binding; Transmembrane;
KW   Multigene family; Vision; Disease mutation; Polymorphism.
FT   TRANSMEM    171    192       POTENTIAL.
FT   TRANSMEM    305    325       POTENTIAL.
FT   TRANSMEM    378    397       POTENTIAL.
FT   NP_BIND     482    605       CAMP.
FT   BINDING     549    549       CAMP (POTENTIAL).
FT   BINDING     564    564       CAMP (POTENTIAL).
FT   VARIANT     153    153       T -> M.
FT                                /FTId=VAR_010902.
FT   VARIANT     163    163       P -> L (IN RMCH).
FT                                /FTId=VAR_010903.
FT   VARIANT     283    283       R -> Q (IN RMCH).
FT                                /FTId=VAR_010904.
FT   VARIANT     283    283       R -> W (IN RMCH).
FT                                /FTId=VAR_010905.
FT   VARIANT     291    291       T -> R (IN RMCH).
FT                                /FTId=VAR_010906.
FT   VARIANT     410    410       R -> W (IN RMCH).
FT                                /FTId=VAR_010910.
FT   VARIANT     529    529       V -> M (IN RMCH).
FT                                /FTId=VAR_010907.
FT   VARIANT     547    547       F -> L (IN RMCH).
FT                                /FTId=VAR_010908.
FT   VARIANT     557    557       G -> R (IN RMCH).
FT                                /FTId=VAR_010909.
```

```
SQ   SEQUENCE   694 AA;  78838 MW;  AE00B4EE760D70A0 CRC64;


  Query Match             51.8%;  Score 1549.5;  DB 1;  Length 694;
  Best Local Similarity   52.6%;  Pred. No. 3.6e-99;
  Matches  290;  Conservative  112;  Mismatches  140;  Indels    9;  Gaps    2;

Qy       9 TTESSPPAPSKARKLLPVLDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHGYLVA 68
           |: ::       :|   |:||| : || ||  :  || ||  :|:|||||  :||  ||:
Db     143 TSNNTEEEKKTKKKDAIVVDPSSNLYYRWLTAIALPVFYNWYLLICRACFDELQSEYLML 202

Qy      69 WLVLDYTSDLLYLLDMVVRFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLASLMPTDVV 128
           ||||||::|:||:||::||  |||||||::| |  |:   |  |  | ||: ||:|||:
Db     203 WLVLDYSADVLYVLDVLVRARTGFLEQGLMVSDTNRLWQHYKTTTQFKLDVLSLVPTDLA 262

Qy     129 YVRLGPHTPTLRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSCLYF 188
           |:::| : | :| || |:  |||| |||||||| ||| |||  |:||| ::|||:|:||
Db     263 YLKVGTNYPEVRFNRLLKFSRLFEFFDRTETRTNYPNMFRIGNLVLYILIIIHWNACIYF 322

Qy     189 ALSRYLGFGRDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAREEEYLFMV 248
           |:|:::||| |:|||: : |   || |:|:|| |:||| |||:|:|||| ::|||||:|
Db     323 AISKFIGFGTDSWVYPNISIPEHGRLSRKYIYSLYWSTLTLTTIGETPPPVKDEEYLFVV 382

Qy     249 GDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVIDWYQ 308
            |||: |: |||||:|:: |:| ||| : | |      :|:||: : | : || ||| |:
Db     383 VDFLVGVLIFATIVGNVGSMISNMNASRAEFQAKIDSIKQYMQFRKVTKDLETRVIRWFD 442

Qy     309 HLQINKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQPQT 368
           :|  |||  :|  :|: ||:::|:||:|::||| || :|:|||:||| || ||||||:|
Db     443 YLWANKKTVDEKEVLKSLPDKLKAEIAINVHLDTLKKVRIFQDCEAGLLVELVLKLRPTV 502

Qy     369 YSPGEYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIINIKGNMSGN 428
           :|||:|:|:|||||:||||| ||:|||||||:||: ||  | |||||||:||||: |||
Db     503 FSPGDYICKKGDIGKEMYIINEGKLAVVADDGVTQFVVLSDGSYFGEISILNIKGSKSGN 562

Qy     429 RRTANIKSLGYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLDVNAEAAEIA 488
           ||||||:|:|||||||||:|| | |:|||:|:  :|||||:||:| | :|     |  |
Db     563 RRTANIRSIGYSDLFCLSKDDLMEALTEYPEAKKALEEKGRQILMKDNLIDEELARAG-A 621

Qy     489 LQEATESRLRGLDQQLDDLQTKFARLLAELESSALKIAYRIERLEWQTR--------EWP 540
             :  | ::  |   || |||:|||||||  :: :|:  |: :|| | :        :
Db     622 DPKDLEEKVEQLGSSLDTLQTRFARLLAEYNATQMKMKQRLSQLESQVKGGGGDKPLADGE 681

Qy     541 MPEDLAEADDE 551
           :| |  : :|:
Db     682 VPGDATKTEDK 692


RESULT   9
CNG2_MOUSE
ID   CNG2_MOUSE     STANDARD;      PRT;   664 AA.
AC   Q62398;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Cyclic-nucleotide-gated olfactory channel (Cyclic-nucleotide-gated
DE   cation channel 2) (CNG channel 2) (CNG-2) (CNG2).
GN   CNGA2 OR CNCG2 OR CNCG4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Heart;
RX   MEDLINE=96439619; PubMed=8841933;
RA   Ruiz M.L., London B., Nadal-Ginard B.;
RT   "Cloning and characterization of an olfactory cyclic nucleotide-gated
RT   channel expressed in mouse heart.";
RL   J. Mol. Cell. Cardiol. 28:1453-1463(1996).
CC   -!- FUNCTION: ODORANT SIGNAL TRANSDUCTION IS PROBABLY MEDIATED
CC       BY A G-PROTEIN COUPLED CASCADE USING CAMP AS SECOND MESSENGER.
CC       THE OLFACTORY CHANNEL CAN BE SHOWN TO BE ACTIVATED BY CYCLIC
CC       NUCLEOTIDES WHICH LEADS TO A DEPOLARIZATION OF OLFACTORY
CC       SENSORY NEURONS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE-GATED CATION CHANNEL
CC       FAMILY.
CC   -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-2 IS THE INITIATOR.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U49391; AAC52712.1; -.
DR   MGD; MGI:108040; Cncg4.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR000595; cNMP_binding.
DR   Pfam; PF00027; cNMP_binding; 1.
DR   Pfam; PF00520; ion_trans; 1.
DR   SMART; SM00100; cNMP; 1.
DR   PROSITE; PS00888; CNMP_BINDING_1; 1.
DR   PROSITE; PS00889; CNMP_BINDING_2; FALSE_NEG.
DR   PROSITE; PS50042; CNMP_BINDING_3; 1.
KW   Ionic channel; Ion transport; cAMP-binding; Transmembrane;
KW   Multigene family; Olfaction.
FT   DOMAIN        1    142       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    143    162       H1 (POTENTIAL).
FT   DOMAIN      164    175       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    176    194       H2 (POTENTIAL).
FT   DOMAIN      195    218       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    219    238       H3 (POTENTIAL).
FT   DOMAIN      239    276       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    277    299       H4 (POTENTIAL).
FT   DOMAIN      300    351       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    352    371       H5 (POTENTIAL).
FT   DOMAIN      372    455       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    456    476       H6 (POTENTIAL).
FT   DOMAIN      477    664       CYTOPLASMIC (POTENTIAL).
FT   NP_BIND     464    586       CAMP (BY SIMILARITY).
FT   BINDING     523    523       CAMP (POTENTIAL).
FT   BINDING     538    538       CAMP (POTENTIAL).
FT   CARBOHYD    381    381       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   664 AA;  76209 MW;  90B2601D727C6AFE CRC64;


  Query Match             51.8%;  Score 1547;  DB 1;  Length 664;
  Best Local Similarity   53.8%;  Pred. No. 5e-99;
  Matches  290;  Conservative  106;  Mismatches  137;  Indels    6;  Gaps    2;

Qy      18 SKARKLLPVLDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHGYLVAWLVLDYTSD 77
           :| :  | ||||:||:|| ||  :  ||:||  :|| |||| |||  | | ||||||
Db     126 TKKKFELFVLDPAGDWYYRWLFVIAMPVLYNWCLLVARACFSDLQRNYFVVWLVLDYFSD 185

Qy      78 LLYLLDMVVRFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLASLMPTDVVYVRLGPHTP 137
             :|: |:::|  |||||||:|| |  ::   |: |  | ||:||::|||::|  :| |:|
Db     186 TVYIADLIIRLRTGFLEQGLLVKDPKKLRDNYIHTLQFKLDVASIIPTDLIYFAVGIHSP 245

Qy     138 TLRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSCLYFALSRYLGFG 197
           :| || |   |:|| |||||||||:||| |||: |:||| |:||||:|:|:|:|: :|||
Db     246 EVRFNRLLHFARMFEFFDRTETRTSYPNIFRISNLVLYILVIIHWNACIYYAISKSIGFG 305

Qy     198 RDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAREEEYLFMVGDFLLAVMG 257
           | ||||:   | :  | |:|:|  |:||| |||:|:|||| ::||||| : |||: |:
Db     306 VDTWVYPNITDPEYGYLAREYIYCLYWSTLTLTTIGETPPPVKDEEYLFFIFDFLIGVLI 365

Qy     258 FATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVIDWYQHLQINKKMT 317
           ||||:|:: |:| |||   | |      || ||: : |:: :| :|| |: :|  |||
Db     366 FATIVGNVGSMISNMNATRAEFQAKIDAVKHYMQFRKVSKDMEAKVIKWFDYLWTNKKTV 425

Qy     318 NEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQPQTYSPGEYVCR 377
           :|  :|::|| :||||:|::|||||| :|:|||:||| || ||||||:|| :|||:|:||
```

```
Db    426 DEREVLKNLPAKLRAEIAINVHLSTLKKVRIFQDCEAGLLVELVLKLRPQVFSPGDYICR 485

Qy    378 KGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIINIKGNMSGNRRTANIKSL 437
          |||||:|||||:||:|||||||||:||||:| ||  ||||||:||||:  |||||  |:||
Db    486 KGDIGKEMYIIKEGKLAVVADDGVTQYALLSAGSCFGEISILNIKGSKMGNRRTGTIRSL 545

Qy    438 GYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLDVNAEAA--EIALQEATES 495
          ||||||||||:|| | ::||| |: ::||:|||||:|   || |  ||  |: :||
Db    546 GYSDLFCLSKDDLMEAVTEYPDAKKVLEERGREILMKEGLLDENEVAASMEVDVQE---- 601

Qy    496 RLRGLDQQLDDLQTKFARLLAELESSALKIAYRIERLEWQTREWPMPEDLAEADDEGEP 554
          :|  |:  :: | |:|||||||   :  |:  ||  || : ::    : |::  :  ||
Db    602 KLEQLETNMETLYTRFARLLAEYTGAQQKLKQRITVLETKMKQNHEDDYLSDGINTPEP 660


RESULT 10
CNG1_BOVIN
ID   CNG1_BOVIN     STANDARD;      PRT;   690 AA.
AC   Q00194;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   cGMP-gated cation channel alpha 1 (CNG channel alpha 1) (CNG-1) (CNG1)
DE   (Cyclic nucleotide gated channel alpha 1) (Cyclic nucleotide gated
DE   channel, photoreceptor) (Cyclic-nucleotide-gated cation channel 1)
DE   (Rod photoreceptor cGMP-gated channel alpha subunit).
GN   CNGA1 OR CNCG1 OR CNCG.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Retinal rod cell;
RX   MEDLINE=90098076; PubMed=2481236;
RA   Kaupp U.B., Niidome T., Tanabe T., Terada S., Boenigk W.,
RA   Stuehmer W., Cook N.J., Kangawa K., Matsuo H., Hirose T., Miyata T.,
RA   Numa S.;
RT   "Primary structure and functional expression from complementary DNA
RT   of the rod photoreceptor cyclic GMP-gated channel.";
RL   Nature 342:762-766(1989).
RN   [2]
RP   3D-STRUCTURE MODELING OF 485-610.
RX   MEDLINE=92256398; PubMed=1316156;
RA   Kumar V.D., Weber I.T.;
RT   "Molecular model of the cyclic GMP-binding domain of the cyclic GMP-
RT   gated ion channel.";
RL   Biochemistry 31:4643-4649(1992).
RN   [3]
RP   TOPOLOGY.
RX   MEDLINE=95365381; PubMed=7543681;
RA   Henn D.K., Baumann A., Kaupp U.B.;
RT   "Probing the transmembrane topology of cyclic nucleotide-gated ion
RT   channels with a gene fusion approach.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:7425-7429(1995).
RN   [4]
RP   TOPOLOGY.
RX   MEDLINE=92112723; PubMed=1370452;
RA   Wohlfart P., Haase W., Molday R.S., Cook N.J.;
RT   "Antibodies against synthetic peptides used to determine the topology
RT   and site of glycosylation of the cGMP-gated channel from bovine rod
RT   photoreceptors.";
RL   J. Biol. Chem. 267:644-648(1992).
CC   -!- FUNCTION: VISUAL SIGNAL TRANSDUCTION IS MEDIATED BY A G-PROTEIN
CC       COUPLED CASCADE USING CGMP AS SECOND MESSENGER. THIS PROTEIN CAN
CC       BE ACTIVATED BY CYCLIC GMP WHICH LEADS TO A OPENING OF THE CATION
CC       CHANNEL AND THEREBY CAUSING A DEPOLARIZATION OF ROD
CC       PHOTORECEPTORS.
CC   -!- SUBUNIT: HOMOTETRAMER OR HIGHER OLIGOMER.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: ROD CELLS IN THE RETINA.
CC   -!- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE-GATED CATION CHANNEL
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X51604; CAA35947.1; -.
DR   PIR; S07103; S07103.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR000595; cNMP_binding.
DR   Pfam; PF00027; cNMP_binding; 1.
DR   Pfam; PF00520; ion_trans; 1.
DR   SMART; SM00100; cNMP; 1.
DR   PROSITE; PS00888; CNMP_BINDING_1; 1.
DR   PROSITE; PS00889; CNMP_BINDING_2; 1.
DR   PROSITE; PS50042; CNMP_BINDING_3; 1.
KW   Ionic channel; Ion transport; cAMP-binding; Transmembrane;
KW   Multigene family; Vision.
FT   DOMAIN        1    162       CYTOPLASMIC (PROBABLE).
FT   TRANSMEM    163    183       H1 (PROBABLE).
FT   DOMAIN      184    196       EXTRACELLULAR (PROBABLE).
FT   TRANSMEM    197    215       H2 (PROBABLE).
FT   DOMAIN      216    239       CYTOPLASMIC (PROBABLE).
FT   TRANSMEM    240    259       H3 (PROBABLE).
FT   DOMAIN      260    297       EXTRACELLULAR (PROBABLE).
FT   TRANSMEM    298    320       H4 (PROBABLE).
FT   DOMAIN      321    372       CYTOPLASMIC (PROBABLE).
FT   TRANSMEM    373    392       H5 (PROBABLE).
FT   DOMAIN      393    476       EXTRACELLULAR (PROBABLE).
FT   TRANSMEM    477    497       H6 (PROBABLE).
FT   DOMAIN      498    690       CYTOPLASMIC (PROBABLE).
FT   NP_BIND     485    607       CGMP (POTENTIAL).
FT   BINDING     544    544       CGMP (POTENTIAL).
FT   BINDING     559    559       CGMP (POTENTIAL).
FT   CARBOHYD    423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   690 AA;  79602 MW;  A01CFB6567424455 CRC64;


  Query Match           51.7%;  Score 1544.5;  DB 1;  Length 690;
  Best Local Similarity 54.0%;  Pred. No. 7.8e-99;
  Matches  285;  Conservative  109;  Mismatches  133;  Indels    1;  Gaps    1;

Qy      6 KVKTTESSPPAPSKARKLLPVLDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHGY 65
            | |  |       : :| : |:|||| || ||  :  |||||  ::: |||| :||  |
Db    135 KKKKEEKGKDKKEEEKKEVVVIDPSGNTYYNWLFCITLPVMYNWTMIIARACFDELQSDY 194

Qy     66 LVAWLVLDYTSDLLYLLDMVVRFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLASLMPT 125
          |  ||  || ||::|||||| ||  ||:||||:|| :: ::  :|  |: | ||: |::||
Db    195 LEYWLAFDYLSDVVYLLDMFVRTRTGYLEQGLLVKEERKLIDKYKSTFQFKLDVLSVIPT 254

Qy    126 DVVYVRLGPHTPTLRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSC 185
          |::|:: | : | :|||| ||  |:|| | |||||| ||| |||: |::|| ::|||||:|
Db    255 DLLYIKFGWNYPEIRLNRLLRISRMFEFFQRTETRTNYPNIFRISNLVMYIIIIIHWNAC 314

Qy    186 LYFALSRYLGFGRDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAREEEYL 245
          :||::|: :||| | |||||   | | || |:|:|| |:||| |||:|:|||| |: ||
Db    315 VYFSISKAIGFGNDTWVYPDVNDPDFGRLARKYVYSLYWSTLTLTTIGETPPPVRDSEYF 374

Qy    246 FMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVID 305
          |:| |||: |: ||||:|:: |:| ||| | | |      :|:||  ::|:: :|:|||
Db    375 FVVADFLIGVLIFATIVGNIGSMISNMNAARAEFQARIDAIKQYMHFRNVSKDMEKRVIK 434

Qy    306 WYQHLQINKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQ 365
          |: :|  |||  :|  :|::||::|||||:|:::||| || :|:|| :||| || |||||||
Db    435 WFDYLWTNKKTVDEREVLKYLPDKLRAEIAINVHLDTLKKVRIFADCEAGLLVELVLKLQ 494

Qy    366 PQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIINIKGNM 425
          || ||||:|:|:|||||:|||||:||:|||||||||||: ||  | |||||||:||||:
Db    495 PQVYSPGDYICKKGDIGREMYIIKEGKLAVVADDGITQFVVLSDGSYFGEISILNIKGSK 554
```

```
Qy     426 SGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLDVNAEAA 485
           :|||||||||||:||||||||||:|| | |:||| |: ::||||::||:|   ||:|   |
Db     555 AGNRRTANIKSIGYSDLFCLSKDDLMEALTEYPDAKGMLEEKGKQILMKDGLLDINIANA 614

Qy     486 EIALQEATESRLRGLDQQLDDLQTKFARLLAELESSALKIAYRIERLE 533
             ::  | ::  ::  :| |||:|||:||| ||   |:  |: ::|
Db     615 GSDPKD-LEEKVTRMESSVDLLQTRFARILAEYESMQQKLKQRLTKVE 661


RESULT 11
CNG1_HUMAN
ID   CNG1_HUMAN     STANDARD;      PRT;   686 AA.
AC   P29973; Q16485; Q16279;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   cGMP-gated cation channel alpha 1 (CNG channel alpha 1) (CNG-1) (CNG1)
DE   (Cyclic nucleotide gated channel alpha 1) (Cyclic nucleotide gated
DE   channel, photoreceptor) (Cyclic-nucleotide-gated cation channel 1)
DE   (Rod photoreceptor cGMP-gated channel alpha subunit).
GN   CNGA1 OR CNCG1 OR CNCG.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Retina;
RX   MEDLINE=92210603; PubMed=1372902;
RA   Pittler S.J., Lee A.K., Altherr M.R., Howard T.A., Seldin M.F.,
RA   Hurwitz R.L., Wasmuth J.J., Baehr W.;
RT   "Primary structure and chromosomal localization of human and mouse
RT   rod photoreceptor cGMP-gated cation channel.";
RL   J. Biol. Chem. 267:6257-6262(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Retina;
RX   MEDLINE=92356211; PubMed=1379636;
RA   Dhallan R.S., Macke J.P., Eddy R.L., Shows T.B., Reed R.R.,
RA   Yau K.-W., Nathans J.;
RT   "Human rod photoreceptor cGMP-gated channel: amino acid sequence,
RT   gene structure, and functional expression.";
RL   J. Neurosci. 12:3248-3256(1992).
RN   [3]
RP   SEQUENCE OF 313-573 FROM N.A.
RX   MEDLINE=95175019; PubMed=7532814;
RA   Distler M., Biel M., Flockerzi V., Hofmann F.;
RT   "Expression of cyclic nucleotide-gated cation channels in non-sensory
RT   tissues and cells.";
RL   Neuropharmacology 33:1275-1282(1994).
RN   [4]
RP   VARIANT ARRP PHE-316, AND VARIANTS GLN-28 AND ASN-114.
RX   MEDLINE=96036047; PubMed=7479749;
RA   Dryja T.P., Finn J.T., Peng Y.-W., McGee T.L., Berson E.L., Yau K.-W.;
RT   "Mutations in the gene encoding the alpha subunit of the rod
RT   cGMP-gated channel in autosomal recessive retinitis pigmentosa.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:10177-10181(1995).
CC   -!- FUNCTION: VISUAL SIGNAL TRANSDUCTION IS MEDIATED BY A G-PROTEIN
CC       COUPLED CASCADE USING CGMP AS SECOND MESSENGER. THIS PROTEIN CAN
CC       BE ACTIVATED BY CYCLIC GMP WHICH LEADS TO A OPENING OF THE CATION
CC       CHANNEL AND THEREBY CAUSING A DEPOLARIZATION OF ROD
CC       PHOTORECEPTORS.
CC   -!- SUBUNIT: HOMOTETRAMER OR HIGHER OLIGOMER. FORMS HETEROOLIGOMERIC
CC       COMPLEX WITH CNG4.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: ROD CELLS IN THE RETINA.
CC   -!- DISEASE: DEFECTS IN CNGA1 ARE A CAUSE OF AUTOSOMAL RECESSIVE
CC       RETINITIS PIGMENTOSA (ARRP). A DISEASE THAT LEADS TO DEGENERATION
CC       OF RETINAL PHOTORECEPTOR CELLS.
CC   -!- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE-GATED CATION CHANNEL
CC       FAMILY.
CC   -!- DATABASE: NAME=Mutations of the CNGA1 gene;
CC       NOTE=Retina International's Scientific Newsletter;
CC       WWW="http://www.retina-international.com/sci-news/cnga1mut.htm".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M84741; AAA52010.1; ALT_INIT.
DR   EMBL; S42457; AAB22778.1; -.
DR   EMBL; S76062; AAD14206.1; -.
DR   PIR; A42161; A42161.
DR   MIM; 123825; -.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR000595; cNMP_binding.
DR   Pfam; PF00027; cNMP_binding; 1.
DR   Pfam; PF00520; ion_trans; 1.
DR   SMART; SM00100; cNMP; 1.
DR   PROSITE; PS00888; CNMP_BINDING_1; 1.
DR   PROSITE; PS00889; CNMP_BINDING_2; 1.
DR   PROSITE; PS50042; CNMP_BINDING_3; 1.
KW   Ionic channel; Ion transport; cAMP-binding; Transmembrane;
KW   Multigene family; Vision; Disease mutation; Polymorphism;
KW   Retinitis pigmentosa.
FT   DOMAIN        1    160       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    161    181       H1 (POTENTIAL).
FT   DOMAIN      182    194       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    195    213       H2 (POTENTIAL).
FT   DOMAIN      214    237       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    238    257       H3 (POTENTIAL).
FT   DOMAIN      258    295       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    296    318       H4 (POTENTIAL).
FT   DOMAIN      319    370       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    371    390       H5 (POTENTIAL).
FT   DOMAIN      391    474       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    475    495       H6 (POTENTIAL).
FT   DOMAIN      496    686       CYTOPLASMIC (POTENTIAL).
FT   NP_BIND     483    605       CGMP (POTENTIAL).
FT   BINDING     542    542       CGMP (POTENTIAL).
FT   BINDING     557    557       CGMP (POTENTIAL).
FT   CARBOHYD    421    421       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT      28     28       R -> Q.
FT                                /FTId=VAR_009295.
FT   VARIANT     114    114       D -> N.
FT                                /FTId=VAR_009296.
FT   VARIANT     316    316       S -> F (IN ARRP).
FT                                /FTId=VAR_009297.
FT   CONFLICT     46     46       S -> Y (IN REF. 1).
FT   CONFLICT     85     85       L -> I (IN REF. 1).
FT   CONFLICT    146    147       EE -> HH (IN REF. 1).
FT   CONFLICT    539    539       Y -> T (IN REF. 1).
FT   CONFLICT    677    678       GA -> WS (IN REF. 1).
SQ   SEQUENCE   686 AA;  79126 MW;  E5200D216FC97AF6 CRC64;


  Query Match             51.5%;  Score 1538.5;  DB 1;  Length 686;
  Best Local Similarity   54.0%;  Pred. No. 2e-98;
  Matches  285;  Conservative  113;  Mismatches  129;  Indels    1;  Gaps    1;

Qy       6 KVKTTESSPPAPSKARKLLPVLDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHGY 65
           | |  | |     : :| : |:||||: || ||  :  |||||  ::: |||| :||  |
Db     133 KKKKEEKSKDKKEEEKKEVVVIDPSGNTYYNWLFCITLPVMYNWTMVIARACFDELQSDY 192

Qy      66 LVAWLVLDYTSDLLYLLDMVVRFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLASLMPT 125
             |  ||:||| ||::||:|| ||  ||:|||||:|| :: :: ::|     | ||: ||:||
Db     193 LEYWLILDYVSDIVYLIDMFVRTRTGYLEQGLLVKEELKLINKYKSNLQFKLDVLSLIPT 252

Qy     126 DVVYVRLGPHTPTLRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSC 185
           |::| :|| : | :|||| ||  |:|| | |||||| ||| |||: |::|| ::|||||:|
Db     253 DLLYFKLGWNYPEIRLNRLLRFSRMFEFFQRTETRTNYPNIFRISNLVMYIVIIIHWNAC 312
```

```
Qy    186 LYFALSRYLGFGRDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAREEEYL 245
          :::::|: :||| | ||||   | | || |:|:|| |:||| |||:|:|||| |: ||:
Db    313 VFYSISKAIGFGNDTWVYPDINDPEFGRLARKYVYSLYWSTLTLTTIGETPPPVRDSEYV 372

Qy    246 FMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVID 305
          |:| |||: |: ||||:|:: |:| ||| | | |      :|:|| ::|:: :|:|||
Db    373 FVVVDFLIGVLIFATIVGNIGSMISNMNAARAEFQARIDAIKQYMHFRNVSKDMEKRVIK 432

Qy    306 WYQHLQINKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQ 365
          |: :|  |||  :|  :|::||::||||:|::|||  || :|:|| :||| || |||||||
Db    433 WFDYLWTNKKTVDEKEVLKYLPDKLRAEIAINVHLDTLKKVRIFADCEAGLLVELVLKLQ 492

Qy    366 PQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIINIKGNM 425
          || ||||:|:|:|||||:|||||:||:|||||||||:||: ||  | |||||||:||||:
Db    493 PQVYSPGDYICKKGDIGREMYIIKEGKLAVVADDGVTQFVVLSDGSYFGEISILNIKGSK 552

Qy    426 SGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLDVNAEAA 485
          :||||||||||:||||||||||:|| | |:||| |:|::||||::||:|   ||:|   |
Db    553 AGNRRTANIKSIGYSDLFCLSKDDLMEALTEYPDAKTMLEEKGKQILMKDGLLDLNIANA 612

Qy    486 EIALQEATESRLRGLDQQLDDLQTKFARLLAELESSALKIAYRIERLE 533
            ::  | ::  ::  :| |||:|||:||| ||   |:  |: ::|
Db    613 GSDPKD-LEEKVTRMEGSVDLLQTRFARILAEYESMQQKLKQRLTKVE 659


RESULT 12
CNG1_MOUSE
ID   CNG1_MOUSE     STANDARD;      PRT;   684 AA.
AC   P29974; Q60776;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   cGMP-gated cation channel alpha 1 (CNG channel alpha 1) (CNG-1) (CNG1)
DE   (Cyclic nucleotide gated channel alpha 1) (Cyclic nucleotide gated
DE   channel, photoreceptor) (Cyclic-nucleotide-gated cation channel 1)
DE   (Rod photoreceptor cGMP-gated channel alpha subunit).
GN   CNGA1 OR CNCG1 OR CNCG.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Retina;
RX   MEDLINE=92210603; PubMed=1372902;
RA   Baehr W., Wasmuth J.J., Hurwitz R.L., Seldin M.F., Howard T.A.,
RA   Altherr M.R., Lee A.K., Pittler S.J.;
RT   "Primary structure and chromosomal localization of human and mouse
RT   rod photoreceptor cGMP-gated cation channel.";
RL   J. Biol. Chem. 267:6257-6262(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RX   MEDLINE=95315239; PubMed=7540868;
RA   Karlson K.H., Ciampolillo-Bates F., McCoy D.E., Kizer N.L.,
RA   Stanton B.A.;
RT   "Cloning of a cGMP-gated cation channel from mouse kidney inner
RT   medullary collecting duct.";
RL   Biochim. Biophys. Acta 1236:197-200(1995).
CC   -!- FUNCTION: VISUAL SIGNAL TRANSDUCTION IS MEDIATED BY A G-PROTEIN
CC       COUPLED CASCADE USING CGMP AS SECOND MESSENGER. THIS PROTEIN CAN
CC       BE ACTIVATED BY CYCLIC GMP WHICH LEADS TO A OPENING OF THE CATION
CC       CHANNEL AND THEREBY CAUSING A DEPOLARIZATION OF ROD
CC       PHOTORECEPTORS.
CC   -!- SUBUNIT: HOMOTETRAMER OR HIGHER OLIGOMER.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: ROD CELLS IN THE RETINA AND INNER MEDULLA OF
CC       KIDNEY.
CC   -!- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE-GATED CATION CHANNEL
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M84742; AAA37425.1; -.
DR   EMBL; U19717; AAA85702.1; -.
DR   EMBL; U19715; AAA85700.1; -.
DR   EMBL; U19716; AAA85701.1; -.
DR   MGD; MGI:88436; Cncg.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR000595; cNMP_binding.
DR   Pfam; PF00027; cNMP_binding; 1.
DR   Pfam; PF00520; ion_trans; 1.
DR   SMART; SM00100; cNMP; 1.
DR   PROSITE; PS00888; CNMP_BINDING_1; 1.
DR   PROSITE; PS00889; CNMP_BINDING_2; 1.
DR   PROSITE; PS50042; CNMP_BINDING_3; 1.
KW   Ionic channel; Ion transport; cAMP-binding; Transmembrane;
KW   Multigene family; Vision.
FT   DOMAIN        1    156       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    157    177       H1 (POTENTIAL).
FT   DOMAIN      178    190       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    191    209       H2 (POTENTIAL).
FT   DOMAIN      210    233       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    234    253       H3 (POTENTIAL).
FT   DOMAIN      254    291       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    292    314       H4 (POTENTIAL).
FT   DOMAIN      315    366       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    367    386       H5 (POTENTIAL).
FT   DOMAIN      387    470       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    471    491       H6 (POTENTIAL).
FT   DOMAIN      492    684       CYTOPLASMIC (POTENTIAL).
FT   NP_BIND     479    601       CGMP (POTENTIAL).
FT   BINDING     538    538       CGMP (POTENTIAL).
FT   BINDING     553    553       CGMP (POTENTIAL).
FT   CARBOHYD    417    417       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    113    114       NK -> I (IN REF. 1).
FT   CONFLICT    200    200       D -> N (IN REF. 1).
FT   CONFLICT    576    576       A -> V (IN REF. 1).
FT   CONFLICT    635    635       R -> C (IN REF. 1).
SQ   SEQUENCE   684 AA;  79460 MW;  13EC2A405B6B8CBF CRC64;


  Query Match           51.3%;  Score 1533.5;  DB 1;  Length 684;
  Best Local Similarity  53.1%;  Pred. No. 4.4e-98;
  Matches  283;  Conservative  112;  Mismatches  127;  Indels   11;  Gaps    2;

Qy      6 KVKTTESSPPAPSKARKLLPVLDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHGY 65
          | |  | :     : :| : |:||||: || ||  :  |||||  ::: |||| :||  |
Db    129 KKKKEEKTKEKKEEEKKEVVVIDPSGNTYYNWLFCITLPVMYNWTMIIARACFDELQSDY 188

Qy     66 LVAWLVLDYTSDLLYLLDMVVRFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLASLMPT 125
          |  ||: || ||::|| || ||  ||:||||:|| |: ::  :|     | ||: |::||
Db    189 LEYWLIFDYVSDVVYLADMFVRTRTGYLEQGLLVKDRMKLIEKYKANLQFKLDVLSVIPT 248

Qy    126 DVVYVRLGPHTPTLRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSC 185
          |::|:: | : | :|||| ||  |:|| | |||||| ||| |||: |::|| ::||||:|
Db    249 DLLYIKFGWNYPEIRLNRLLRISRMFEFFQRTETRTNYPNIFRISNLVMYIVIIIHWNAC 308

Qy    186 LYFALSRYLGFGRDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAREEEYL 245
          :|:::|: :||| | |||||   | | || |:|:|| |:||| |||:|:|||| : ||:
Db    309 VYYSISKAIGFGNDTWVYPDVNDPEFGRLARKYVYSLYWSTLTLTTIGETPPPVLDSEYI 368

Qy    246 FMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVID 305
          |:| |||: |: ||||:|:: |:| ||| | | |      :|:||  ::|:: :|:|||
Db    369 FVVVDFLIGVLIFATIVGNIGSMISNMNAARAEFQSRVDAIKQYMNFRNVSKDMEKRVIK 428

Qy    306 WYQHLQINKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQ 365
          |: :|  |||  :|  :|::||::||||:|::||| || :|:|| :||| || |||||||
Db    429 WFDYLWTNKKTVDEREVLRYLPDKLRAEIAINVHLDTLKKVRIFADCEAGLLVELVLKLQ 488
```

```
Qy    366 PQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIINIKGNM 425
          || ||||:|:|:|||||:|||||:||:||||||||||||: ||  | |||||||:||||:
Db    489 PQVYSPGDYICKKGDIGREMYIIKEGKLAVVADDGITQFVVLSDGSYFGEISILNIKGSK 548

Qy    426 SGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLDVN---- 481
          :|||||||||:|||||||||:|| | |:||| |:|::|||||:||:|   ||:|
Db    549 AGNRRTANIKSIGYSDLFCLSKDDLMEALTEYPDAKTMLEEKGRQILMKDGLLDINIANM 608

Qy    482 -AEAAEIALQEATESRLRGLDQQLDDLQTKFARLLAELESSALKIAYRIERLE 533
           ::  ::      | ::  ::  :| |||:|||:||| ||   |:  |: ::|
Db    609 GSDPKDL------EEKVTRMEGSVDLLQTRFARILAEYESMQQKLKQRLTKVE 655


RESULT 13
CNG1_CANFA
ID   CNG1_CANFA     STANDARD;      PRT;   691 AA.
AC   Q28279;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   cGMP-gated cation channel alpha 1 (CNG channel alpha 1) (CNG-1) (CNG1)
DE   (Cyclic nucleotide gated channel alpha 1) (Cyclic nucleotide gated
DE   channel, photoreceptor) (Cyclic-nucleotide-gated cation channel 1)
DE   (Rod photoreceptor cGMP-gated channel alpha subunit).
GN   CNGA1 OR CNCG1 OR CNCG.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BEAGLE X BRIARD;
RA   Veske A.;
RL   Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97415722; PubMed=9268598;
RA   Zhang Q., Pearce-Kelling S., Acland G.M., Aguirre G.D., Ray K.;
RT   "Canine rod photoreceptor cGMP-gated channel protein alpha-subunit:
RT   studies on the expression of the gene and characterization of the
RT   cDNA.";
RL   Exp. Eye Res. 65:301-309(1997).
CC   -!- FUNCTION: VISUAL SIGNAL TRANSDUCTION IS MEDIATED BY A G-PROTEIN
CC       COUPLED CASCADE USING CGMP AS SECOND MESSENGER. THIS PROTEIN CAN
CC       BE ACTIVATED BY CYCLIC GMP WHICH LEADS TO A OPENING OF THE CATION
CC       CHANNEL AND THEREBY CAUSING A DEPOLARIZATION OF ROD
CC       PHOTORECEPTORS.
CC   -!- SUBUNIT: HOMOTETRAMER OR HIGHER OLIGOMER.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE-GATED CATION CHANNEL
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X99914; CAA68186.1; -.
DR   EMBL; U83905; AAB61707.1; -.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR000595; cNMP_binding.
DR   Pfam; PF00027; cNMP_binding; 1.
DR   Pfam; PF00520; ion_trans; 1.
DR   SMART; SM00100; cNMP; 1.
DR   PROSITE; PS00888; CNMP_BINDING_1; 1.
DR   PROSITE; PS00889; CNMP_BINDING_2; 1.
DR   PROSITE; PS50042; CNMP_BINDING_3; 1.
KW   Ionic channel; Ion transport; cAMP-binding; Transmembrane;
KW   Multigene family; Vision.

FT   DOMAIN        1    163       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    164    184       H1 (POTENTIAL).
FT   DOMAIN      185    197       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    198    216       H2 (POTENTIAL).
FT   DOMAIN      217    240       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    241    260       H3 (POTENTIAL).
FT   DOMAIN      261    298       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    299    321       H4 (POTENTIAL).
FT   DOMAIN      322    373       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    374    393       H5 (POTENTIAL).
FT   DOMAIN      394    477       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    478    498       H6 (POTENTIAL).
FT   DOMAIN      499    691       CYTOPLASMIC (POTENTIAL).
FT   NP_BIND     486    608       CGMP (POTENTIAL).
FT   BINDING     545    545       CGMP (POTENTIAL).
FT   BINDING     560    560       CGMP (POTENTIAL).
FT   CARBOHYD    424    424       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   691 AA;  80251 MW;  0775CAA42F065275 CRC64;


  Query Match           51.3%;  Score 1532.5;  DB 1;  Length 691;
  Best Local Similarity  53.6%;  Pred. No. 5.2e-98;
  Matches  283;  Conservative  110;  Mismatches  134;  Indels    1;  Gaps    1;

Qy      6 KVKTTESSPPAPSKARKLLPVLDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHGY 65
          |    |       : :| : |:||:|: || ||  :  |||||  ::: |||| :|| |
Db    136 KKNKEEKGKDKKEEEKKEVMVIDPAGNMYYNWLFCITLPVMYNWTMVIARACFDELQSDY 195

Qy     66 LVAWLVLDYTSDLLYLLDMVVRFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLASLMPT 125
          |  |:: || ||::|||||  ||  ||:||||:|| :: ::  :|      | ||  |::||
Db    196 LEYWIIFDYLSDIVYLLDMFVRTRTGYLEQGLLVREEAKLIEKYKSNLQFKLDFLSVIPT 255

Qy    126 DVVYVRLGPHTPTLRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSC 185
          |::| :|| : | :|||| ||  |:|| | |||||| ||| |||: |::|| ::||||:|
Db    256 DLLYFKLGWNYPEIRLNRLLRISRMFEFFQRTETRTNYPNIFRISNLVMYIVIIIHWNAC 315

Qy    186 LYFALSRYLGFGRDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAREEEYL 245
          :||::|: :||| | |||||   | | || |:|:|| |:||| |||:|:|||| |: ||:
Db    316 VYFSISKAIGFGNDTWVYPDVNDPEFGRLARKYVYSLYWSTLTLTTIGETPPPVRDSEYV 375

Qy    246 FMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVID 305
          |:| |||: |: ||||:|:: |:| ||| | | |      :|:||  ::|:: :|:|||
Db    376 FVVVDFLIGVLIFATIVGNIGSMISNMNAARAEFQARIDAIKQYMHFRNVSKDMEKRVIK 435

Qy    306 WYQHLQINKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQ 365
          |: :|  |||  :|  :|::||::||||:|::||| || :|:|| :||| || |||||||
Db    436 WFDYLWTNKKTVDEKEVLKYLPDKLRAEIAINVHLDTLKKVRIFADCEAGLLVELVLKLQ 495

Qy    366 PQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIINIKGNM 425
          || ||||:|:|:|||||:|||||:||:||||||||||||: ||  | |||||||:||||:
Db    496 PQVYSPGDYICKKGDIGREMYIIKEGKLAVVADDGITQFVVLSDGSYFGEISILNIKGSK 555

Qy    426 SGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLDVNAEAA 485
          :|||||||||:|||||||||:|| | |:||| |:|::||||::||:|   ||:|   |
Db    556 AGNRRTANIKSIGYSDLFCLSKDDLMEALTEYPDAKTMLEEKGKQILMKDGLLDINIANA 615

Qy    486 EIALQEATESRLRGLDQQLDDLQTKFARLLAELESSALKIAYRIERLE 533
           ::  | ::  ::  :| |||:|||:||| ||   |:  |: ::|
Db    616 GSDPKD-LEEKVTRMEGSVDLLQTRFARILAEYESMQQKLKQRLTKVE 662


RESULT 14
CNG1_RAT
ID   CNG1_RAT       STANDARD;      PRT;   683 AA.
AC   Q62927; O08659;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   cGMP-gated cation channel alpha 1 (CNG channel alpha 1) (CNG-1) (CNG1)
DE   (Cyclic nucleotide gated channel alpha 1) (Cyclic nucleotide gated
DE   channel, photoreceptor) (Cyclic-nucleotide-gated cation channel 1)
DE   (Rod photoreceptor cGMP-gated channel alpha subunit).
GN   CNGA1 OR CNCG1 OR CNCG.
```

```
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY;
RA   Barnstable C.J., Wei J.Y.;
RL   Submitted (MAR-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97287732; PubMed=9142860;
RA   Ding C., Potter E.D., Qiu W., Coon S.L., Levine M.A., Guggino S.E.;
RT   "Cloning and widespread distribution of the rat rod-type cyclic
RT   nucleotide-gated cation channel.";
RL   Am. J. Physiol. 272:C1335-C1344(1997).
RN   [3]
RP   SEQUENCE OF 521-683 FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY;
RX   MEDLINE=97197878; PubMed=9045728;
RA   Bradley J., Zhang Y., Bakin R., Lester H.A., Ronnett G.V., Zinn K.;
RT   "Functional expression of the heteromeric 'olfactory' cyclic
RT   nucleotide- gated channel in the hippocampus: a potential effector of
RT   synaptic plasticity in brain neurons.";
RL   J. Neurosci. 17:1993-2005(1997).
CC   -!- FUNCTION: VISUAL SIGNAL TRANSDUCTION IS MEDIATED BY A G-PROTEIN
CC       COUPLED CASCADE USING CGMP AS SECOND MESSENGER. THIS PROTEIN CAN
CC       BE ACTIVATED BY CYCLIC GMP WHICH LEADS TO A OPENING OF THE CATION
CC       CHANNEL AND THEREBY CAUSING A DEPOLARIZATION OF ROD
CC       PHOTORECEPTORS.
CC   -!- SUBUNIT: HOMOTETRAMER OR HIGHER OLIGOMER.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: ROD CELLS IN THE RETINA.
CC   -!- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE-GATED CATION CHANNEL
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U48803; AAA92110.1; -.
DR   EMBL; U93851; AAC53139.1; -.
DR   EMBL; U76220; AAC17594.1; -.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR000595; cNMP_binding.
DR   Pfam; PF00027; cNMP_binding; 1.
DR   Pfam; PF00520; ion_trans; 1.
DR   SMART; SM00100; cNMP; 1.
DR   PROSITE; PS00888; CNMP_BINDING_1; 1.
DR   PROSITE; PS00889; CNMP_BINDING_2; 1.
DR   PROSITE; PS50042; CNMP_BINDING_3; 1.
KW   Ionic channel; Ion transport; cAMP-binding; Transmembrane;
KW   Multigene family; Vision.
FT   DOMAIN       1    155       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   156    176       H1 (POTENTIAL).
FT   DOMAIN     177    189       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   190    208       H2 (POTENTIAL).
FT   DOMAIN     209    232       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   233    252       H3 (POTENTIAL).
FT   DOMAIN     253    290       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   291    313       H4 (POTENTIAL).
FT   DOMAIN     314    365       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   366    385       H5 (POTENTIAL).
FT   DOMAIN     386    469       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   470    490       H6 (POTENTIAL).
FT   DOMAIN     491    683       CYTOPLASMIC (POTENTIAL).
FT   NP_BIND    478    600       CGMP (POTENTIAL).
FT   BINDING    537    537       CGMP (POTENTIAL).
FT   BINDING    552    552       CGMP (POTENTIAL).
FT   CARBOHYD   416    416       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     3      4       TN -> KV (IN REF. 2).
FT   CONFLICT   141    141       E -> G (IN REF. 2).
FT   CONFLICT   224    224       E -> K (IN REF. 2).
SQ   SEQUENCE   683 AA;  79228 MW;  62D49F64F8D917C5 CRC64;


  Query Match             51.3%;  Score 1532;  DB 1;  Length 683;
  Best Local Similarity   52.6%;  Pred. No. 5.6e-98;
  Matches  293;  Conservative  110;  Mismatches  144;  Indels   10;  Gaps    5;

Qy       6 KVKTTESSP-PAPSKARKLLPVLDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHG 64
           | |  |  |     : :| : |:|||||: || ||  :  |||||  ::: |||| :||
Db     127 KEKKKEEKPKEKKEEEKKEVVVIDPSGNMYYNWLFCITLPVMYNWTMIIARACFDELQSD 186

Qy      65 YLVAWLVLDYTSDLLYLLDMVVRFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLASLMP 124
           ||  ||: || ||::|| || ||  ||:||||:|| |: ::  :|     | ||: |::|
Db     187 YLEYWLIFDYVSDVVYLADMFVRTRTGYLEQGLLVKDELKLIEKYKANLQFKLDVLSVIP 246

Qy     125 TDVVYVRLGPHTPTLRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNS 184
           ||::| : | : | :|||| ||  |:|| | |||||| ||| |||: |::|| ::|||||:
Db     247 TDLLYFKFGWNYPEIRLNRLLRISRMFEFFQRTETRTNYPNIFRISNLVMYIVIIIHWNA 306

Qy     185 CLYFALSRYLGFGRDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAREEEY 244
           |:|:::|: :||| | |||||   | | || |:|:|| |:||| |||:|:|||| : ||
Db     307 CVYYSISKAIGFGNDTWVYPDVNDPEFGRLARKYVYSLYWSTLTLTTIGETPPPVLDSEY 366

Qy     245 LFMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVI 304
           :|:| |||: |: ||||:|:: |:| ||| | | |       :|:||  ::|:: :|:|||
Db     367 VFVVVDFLIGVLIFATIVGNIGSMISNMNAARAEFQSRVDAIKQYMNFRNVSKDMEKRVI 426

Qy     305 DWYQHLQINKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKL 364
            |: :|  |||  :|  :|::||::||||:|::||| || :|:|| :||| || ||||||
Db     427 KWFDYLWTNKKTVDEREVLRYLPDKLRAEIAINVHLDTLKKVRIFADCEAGLLVELVLKL 486

Qy     365 QPQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIINIKGN 424
           ||| ||||:|:|:|||||:|||||:||:|||||||||||: ||  | |||||||:||||:
Db     487 QPQVYSPGDYICKKGDIGREMYIIKEGKLAVVADDGITQFVVLSDGSYFGEISILNIKGS 546

Qy     425 MSGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLDVNAEA 484
           :|||||||||:|||||||||:|| | |:||| |:|::|||||:||:|   ||:|
Db     547 KAGNRRTANIKSIGYSDLFCLSKDDLMEALTEYPDAKTMLEEKGRQILMKDGLLDIN--I 604

Qy     485 AEIALQ-EATESRLRGLDQQLDDLQTKFARLLAELESSALKIAYRIERLEWQTREWPMPE 543
           | :   :  | ::  ::  :| |||:|||:||| ||   |:  |: ::|   :  |: |
Db     605 ANLGSDPKDLEEKVTRMEGSVDLLQTRFARILAEYESMQQKLKQRLTKVEKFLK--PLIE 662

Qy     544 D----LAEADDEGEPEE 556
               | |   | || |
Db     663 TEFSALEEPGGESEPTE 679


RESULT  15
CNG3_CHICK
ID   CNG3_CHICK     STANDARD;      PRT;   645 AA.
AC   Q90980;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Cyclic nucleotide gated channel, ROD photoreceptor, alpha subunit
DE   (CNG channel 3) (CNG-3) (CNG3).
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93264082; PubMed=7684234;
RA   Boenigk W., Altenhofen W., Mueller F., Dose A., Illing M.,
RA   Molday R.S., Kaupp U.B.;
RT   "Rod and cone photoreceptor cells express distinct genes for
RT   cGMP-gated channels.";
```

```
RL   Neuron 10:865-877(1993).
CC   -!- FUNCTION: VISUAL SIGNAL TRANSDUCTION IS MEDIATED BY A G-PROTEIN
CC       COUPLED CASCADE USING CGMP AS SECOND MESSENGER. THIS PROTEIN CAN
CC       BE ACTIVATED BY CYCLIC GMP WHICH LEADS TO A OPENING OF THE CATION
CC       CHANNEL AND THEREBY CAUSING A DEPOLARIZATION OF ROD
CC       PHOTORECEPTORS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE-GATED CATION CHANNEL
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X89599; CAA61758.1; -.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR000595; cNMP_binding.
DR   Pfam; PF00027; cNMP_binding; 1.
DR   Pfam; PF00520; ion_trans; 1.
DR   SMART; SM00100; cNMP; 1.
DR   PROSITE; PS00888; CNMP_BINDING_1; 1.
DR   PROSITE; PS00889; CNMP_BINDING_2; 1.
DR   PROSITE; PS50042; CNMP_BINDING_3; 1.
KW   Ionic channel; Ion transport; cAMP-binding; Transmembrane; Vision;
KW   Multigene family.
FT   DOMAIN        1    117       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    118    137       H1 (POTENTIAL).
FT   DOMAIN      138    150       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    151    169       H2 (POTENTIAL).
FT   DOMAIN      170    193       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    194    213       H3 (POTENTIAL).
FT   DOMAIN      214    251       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    252    274       H4 (POTENTIAL).
FT   DOMAIN      275    326       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    327    346       H5 (POTENTIAL).
FT   DOMAIN      347    430       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    431    451       H6 (POTENTIAL).
FT   DOMAIN      452    645       CYTOPLASMIC (POTENTIAL).
FT   NP_BIND     439    561       CAMP (BY SIMILARITY).
FT   BINDING     498    498       CAMP (POTENTIAL).
FT   BINDING     513    513       CAMP (POTENTIAL).
FT   CARBOHYD    377    377       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   645 AA;  74778 MW;  989D515F61AC7D31 CRC64;
```

```
  Query Match             51.2%;  Score 1529;  DB 1;  Length 645;
  Best Local Similarity   50.7%;  Pred. No. 8.3e-98;
  Matches  285;  Conservative  124;  Mismatches  143;  Indels   10;  Gaps    5;

Qy       6 KVKTTESSPPAPSKARKLLPVLDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHGY 65
           | |  :       : :| : ::||:|: || ||  :  |||||  ::: |||| :||: |
Db      89 KHKNKDKKKGKEEEKKKDIFIIDPAGNMYYNWLFCITMPVMYNWTMIIARACFDELQNDY 148

Qy      66 LVAWLVLDYTSDLLYLLDMVVRFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLASLMPT 125
           |  | ::|| ||::|: || ||  ||:||||:|| :: ::  :|  :  | ||  |::||
Db     149 LAVWFIVDYVSDVIYIADMFVRTRTGYLEQGLLVKEEQKLKEKYKSSLQFKLDFLSIIPT 208

Qy     126 DVVYVRLGPHTPTLRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSC 185
           |::| :|| : | ||:|| ||  |:|| | |||||| ||| |||: |::|| ::||||:|
Db     209 DLLYFKLGLNYPELRINRLLRVARMFEFFQRTETRTNYPNIFRISNLVMYIVIIIHWNAC 268

Qy     186 LYFALSRYLGFGRDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAREEEYL 245
           :|:::|: :||| | ||||: : | | || |:|:|| |:||| |||:|:|||| |: ||
Db     269 VYYSISKAIGFGADTWVYPNTSHPEFARLTRKYVYSLYWSTLTLTTIGETPPPVRDSEYF 328

Qy     246 FMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVID 305
           |:| |||: |: ||||:|:: |:| ||| | | |      :|:||  ::|:: :|:|||
Db     329 FVVVDFLVGVLIFATIVGNVGSMISNMNAARAEFQAKIDAIKQYMHFRNVSKDMEKRVIK 388

Qy     306 WYQHLQINKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQ 365
           |: :|  |||  :|  :|::||::||||:|::||| || :|:|| :||| || |||||||
Db     389 WFDYLWTNKKAVDEREVLKYLPDKLRAEIAINVHLETLKKVRIFADCEAGLLVELVLKLQ 448

Qy     366 PQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIINIKGNM 425
           || ||||:|:|||||||:|||||:||:|||||||:||: ||  | |||||||:||||:
Db     449 PQVYSPGDYICRKGDIGREMYIIKEGKLAVVADDGVTQFVVLSDGSYFGEISILNIKGSK 508

Qy     426 SGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLDVNAEAA 485
           :|||||||:|:|||||||||:|| | |:||| |: ::||||::||:|   ||:  | |
Db     509 AGNRRTANIRSIGYSDLFCLSKDDLMEALTEYPDAKAMLEEKGKQILMKDGLLDI--EVA 566

Qy     486 EIALQ-EATESRLRGLDQQLDDLQTKFARLLAELESSALKIAYRIERLEWQTREWPMPE- 543
           :    :  | ::  ::  :| |||||||||| :::  |:  |: ::|   :  |: |
Db     567 NLGSDPKDLEEKVAYMEGSMDRLQTKFARLLAEYDAAQQKLKKRLTQIEKILK--PVMEQ 624

Qy     544 ---DLAEADDEGEPEEGTSKDE 562
              |  |||   : : | :| |
Db     625 EFLDFEEADPPTD-KPGVTKTE 645
```

Search completed: June 20, 2002, 17:15:51
Job time: 534 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 20, 2002, 17:06:12 ; Search time 86.97 Seconds
(without alignments)
1143.752 Million cell updates/sec

Title: US-09-927-267-1
Perfect score: 2989
Sequence: 1 MSQDTKVKTTESSPPAPSKA.....EGTSKDEEGRASQEGPPGPE 575

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1578	52.8	611	11	Q9QWN7	Q9qwn7 rattus norv
2	1578	52.8	632	11	Q9ER32	Q9er32 rattus norv
3	1578	52.8	670	11	Q9ER33	Q9er33 rattus norv
4	1538.5	51.5	686	6	Q9N0H4	Q9n0h4 sus scrofa
5	1119.5	37.5	733	5	P90800	P90800 caenorhabdi
6	990.5	33.1	905	5	O97119	O97119 limulus pol
7	947	31.7	702	5	Q9W202	Q9w202 drosophila
8	879	29.4	1463	5	Q9U5E2	Q9u5e2 drosophila
9	829.5	27.8	1693	5	Q9VXV8	Q9vxv8 drosophila
10	829	27.7	252	11	Q9QX26	Q9qx26 rattus norv
11	788.5	26.4	800	5	P90975	P90975 caenorhabdi
12	788.5	26.4	800	5	O62237	O62237 caenorhabdi
13	781.5	26.1	1037	5	Q9W2D5	Q9w2d5 drosophila
14	748	25.0	695	5	Q93486	Q93486 caenorhabdi
15	738	24.7	938	6	O77658	O77658 bos taurus
16	738	24.7	948	6	O77659	O77659 bos taurus
17	738	24.7	952	6	O77660	O77660 bos taurus
18	732	24.5	858	11	O55157	O55157 rattus norv
19	732	24.5	1339	11	O35788	O35788 rattus norv
20	719	24.1	1251	4	Q9UMG2	Q9umg2 homo sapien
21	716	24.0	1245	4	O43636	O43636 homo sapien
22	681	22.8	515	5	Q95SL0	Q95sl0 drosophila
23	667	22.3	832	5	Q9N4C1	Q9n4c1 caenorhabdi
24	655	21.9	809	4	Q9NQW8	Q9nqw8 homo sapien
25	647	21.6	644	5	O61827	O61827 caenorhabdi
26	627.5	21.0	610	4	Q9NRE9	Q9nre9 homo sapien
27	620	20.7	694	11	Q9JJZ9	Q9jjz9 mus musculu
28	551	18.4	151	11	P70607	P70607 rattus norv
29	498.5	16.7	678	5	O96777	O96777 heliothis v
30	488	16.3	418	5	Q9W201	Q9w201 drosophila
31	477	16.0	945	5	Q9Y1J9	Q9y1j9 drosophila
32	470.5	15.7	767	5	Q20170	Q20170 caenorhabdi
33	468.5	15.7	1198	11	Q9JKA7	Q9jka7 rattus norv
34	467.5	15.6	1203	4	Q9Y3Q4	Q9y3q4 homo sapien
35	467.5	15.6	1203	4	Q9UMQ7	Q9umq7 homo sapien
36	465.5	15.6	889	4	Q9UBS2	Q9ubs2 homo sapien
37	465.5	15.6	889	4	Q9UL51	Q9ul51 homo sapien
38	464.5	15.5	863	11	O88703	O88703 mus musculu
39	463.5	15.5	834	11	Q9JKA9	Q9jka9 rattus norv
40	463.5	15.5	1175	6	Q9TV66	Q9tv66 oryctolagus
41	457.5	15.3	504	11	O70506	O70506 mus musculu
42	456.5	15.3	597	4	O60742	O60742 homo sapien
43	456.5	15.3	703	4	Q9P1Z3	Q9plz3 homo sapien
44	456.5	15.3	779	11	O88705	O88705 mus musculu
45	456.5	15.3	780	11	Q9JKA8	Q9jka8 rattus norv

ALIGNMENTS

RESULT 1
Q9QWN7
ID Q9QWN7 PRELIMINARY; PRT; 611 AA.
AC Q9QWN7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE CYCLIC NUCLEOTIDE-GATED CHANNEL.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97426406; PubMed=9278419;
RA Misaka T., Kusakabe Y., Emori Y., Gonoi T., Arai S., Abe K.;
RT "Taste buds have a cyclic nucleotide-activated channel, CNGgust.";
RL J. Biol. Chem. 272:22623-22629(1997).
DR EMBL; AB002801; BAA24353.1; -.
DR InterPro; IPR000636; Cation_chan_non_lig.
DR InterPro; IPR001622; Channel_pore_K.
DR InterPro; IPR000595; cNMP_binding.
DR Pfam; PF00027; cNMP_binding; 1.
DR Pfam; PF00520; ion_trans; 1.
DR SMART; SM00100; cNMP; 1.
DR PROSITE; PS00888; CNMP_BINDING_1; 1.
DR PROSITE; PS00889; CNMP_BINDING_2; 1.
DR PROSITE; PS50042; CNMP_BINDING_3; 1.
SQ SEQUENCE 611 AA; 70399 MW; 9ACECEF9EEF63AC4 CRC64;

Query Match 52.8%; Score 1578; DB 11; Length 611;
Best Local Similarity 54.5%; Pred. No. 6.1e-110;
Matches 300; Conservative 105; Mismatches 125; Indels 20; Gaps 5;

```
Qy      14 PPAPSKARKLLP-VLDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHGYLVAWLVL 72
           ||   | ||  | |:||| : || ||  :  || ||  :|||||||:||  :|  |'|||
Db      68 PPEGGKGRKKDPIVVDPSSNIYYRWLTAIALPVFYNWCLLVCRACFDELQSEHLTLWLVL 127
```

```
Qy      73 DYTSDLLYLLDMVVRFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLASLMPTDVVYVRL 132
           ||::| ||::||:|| |||||||::| |  |:   | :|  | ||: ||:|||: |::|
Db     128 DYSADALYVVDMLVRARTGFLEQGLMVRDTKRLWKHYTKTLHFKLDILSLIPTDLAYLKL 187

Qy     133 GPHTPTLRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSCLYFALSR 192
           | : | || || ||  |||| |||||||| ||| |||  |:||  ::||||:|:|||:|:
Db     188 GMNYPELRFNRLLRFSRLFEFFDRTETRTNYPNVFRIGNLVLYTLIIIHWNACIYFAISK 247

Qy     193 YLGFGRDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAREEEYLFMVGDFL 252
           ::||| |:|||: ::| : || |:|:|| |:||| |||:|:|||| ::|||||:| |||
Db     248 FIGFGTDSWVYPNTSKPEYGRLSRKYIYSLYWSTLTLTTIGETPPPVKDEEYLFVVIDFL 307

Qy     253 LAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVIDWYQHLQI 312
           : |: ||||:|:: |:| ||| : | |      :|:||: : | : || ||| |: :|
Db     308 VGVLIFATIVGNVGSMISNMNASRAEFQAKIDSIKQYMQFRKVTKDLETRVIRWFDYLWA 367

Qy     313 NKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQPQTYSPG 372
           |:|  :|  :|::||:::|:||:|::||| || :|:|||:||| || ||||||:|  :|||
Db     368 NRKTVDEKEVLKNLPDKLKAEIAINVHLDTLKKVRIFQDCEAGLLVELVLKLRPAVFSPG 427

Qy     373 EYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIINIKGNMSGNRRTA 432
           :|:|:|||||:|||||:||:||||||||:|| ||  | |||||||:|||: |||||||
Db     428 DYICKKGDIGREMYIIKEGKLAVVADDGVTQFVVLSDGSYFGEISILNIKGSKSGNRRTA 487

Qy     433 NIKSLGYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLD-----VNAEAAEI 487
           ||:|:|||||||||:|| | |:||| |:  :|||||:||:| | :|       |:|  |
Db     488 NIRSIGYSDLFCLSKDDLMETLTEYPDAKRALEEKGRQILMKDNLIDDDLVTARADARNI 547

Qy     488 ALQEATESRLRGLDQQLDDLQTKFARLLAELESSALKIAYRIERLEWQ-TR-------EW 539
                 | ::  |:  || |||:|||||||  :| :|:  |: :|| | ||       :
Db     548 ------EEKVEYLESSLDGLQTRFARLLAEYSASQMKLKQRLSQLESQMTRRGHGFSPDR 601

Qy     540 PMPEDLAEAD 549
              || ::||
Db     602 ENSEDASKAD 611


RESULT   2
Q9ER32
ID   Q9ER32      PRELIMINARY;      PRT;   632 AA.
AC   Q9ER32;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CYCLIC NUCLEOTIDE-GATED CHANNEL 2A.
CS   Rattus norvegicus (Rat).
CC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=RETINA;
RX   MEDLINE=20442421; PubMed=10984544;
RA   Meyer M.R., Angele A., Kremmer E., Kaupp U.B., Mueller F.;
RT   "A cGMP-signalling pathway in a subset of olfactory sensory neurons.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:10595-10600(2000).
DR   EMBL; AJ272429; CAC09431.1; -.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR001622; Channel_pore_K.
DR   InterPro; IPR000595; cNMP_binding.
DR   Pfam; PF00027; cNMP_binding; 1.
DR   Pfam; PF00520; ion_trans; 1.
DR   SMART; SM00100; cNMP; 1.
DR   PROSITE; PS00888; CNMP_BINDING_1; 1.
DR   PROSITE; PS50042; CNMP_BINDING_3; 1.
SQ   SEQUENCE   632 AA;  72536 MW;  2A48AAECDD600D37 CRC64;


  Query Match             52.8%;  Score 1578;  DB 11;  Length 632;
  Best Local Similarity   54.5%;  Pred. No. 6.4e-110;
  Matches  300;  Conservative  105;  Mismatches  125;  Indels   20;  Gaps    5;

Qy      14 PPAPSKARKLLP-VLDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHGYLVAWLVL 72
           ||   | ||  | |:||| : || ||  :  || ||  :|||||| :||  :|  ||||
Db      89 PPEGGKGRKKDPIVVDPSSNIYYRWLTAIALPVFYNWCLLVCRACFDELQSEHLTLWLVL 148

Qy      73 DYTSDLLYLLDMVVRFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLASLMPTDVVYVRL 132
           ||::| ||::||:|| |||||||::| |  |:   | :|  | ||: ||:|||: |::|
Db     149 DYSADALYVVDMLVRARTGFLEQGLMVRDTKRLWKHYTKTLHFKLDILSLIPTDLAYLKL 208

Qy     133 GPHTPTLRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSCLYFALSR 192
           | : | || || ||  |||| |||||||| ||| |||  |:||  ::||||:|:|||:|:
Db     209 GMNYPELRFNRLLRFSRLFEFFDRTETRTNYPNVFRIGNLVLYTLIIIHWNACIYFAISK 268

Qy     193 YLGFGRDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAREEEYLFMVGDFL 252
           ::||| |:|||: ::| : || |:|:|| |:||| |||:|:|||| ::|||||:| |||
Db     269 FIGFGTDSWVYPNTSKPEYGRLSRKYIYSLYWSTLTLTTIGETPPPVKDEEYLFVVIDFL 328

Qy     253 LAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVIDWYQHLQI 312
           : |: ||||:|:: |:| ||| : | |      :|:||: : | : || ||| |: :|
Db     329 VGVLIFATIVGNVGSMISNMNASRAEFQAKIDSIKQYMQFRKVTKDLETRVIRWFDYLWA 388

Qy     313 NKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQPQTYSPG 372
           |:|  :|  :|::||:::|:||:|::||| || :|:|||:||| || ||||||:|  :|||
Db     389 NRKTVDEKEVLKNLPDKLKAEIAINVHLDTLKKVRIFQDCEAGLLVELVLKLRPAVFSPG 448

Qy     373 EYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIINIKGNMSGNRRTA 432
           :|:|:|||||:|||||:||:||||||||:|| ||  | |||||||:|||: |||||||
Db     449 DYICKKGDIGREMYIIKEGKLAVVADDGVTQFVVLSDGSYFGEISILNIKGSKSGNRRTA 508

Qy     433 NIKSLGYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLD-----VNAEAAEI 487
           ||:|:|||||||||:|| | |:||| |:  :|||||:||:| | :|       |:|  |
Db     509 NIRSIGYSDLFCLSKDDLMETLTEYPDAKRALEEKGRQILMKDNLIDDDLVTARADARNI 568

Qy     488 ALQEATESRLRGLDQQLDDLQTKFARLLAELESSALKIAYRIERLEWQ-TR-------EW 539
                 | ::  |:  || |||:|||||||  :| :|:  |: :|| | ||       :
Db     569 ------EEKVEYLESSLDGLQTRFARLLAEYSASQMKLKQRLSQLESQMTRRGHGFSPDR 622

Qy     540 PMPEDLAEAD 549
              || ::||
Db     623 ENSEDASKAD 632


RESULT   3
Q9ER33
ID   Q9ER33      PRELIMINARY;      PRT;   670 AA.
AC   Q9ER33;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   CYCLIC NUCLEOTIDE-GATED CHANNEL 2B.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=OLFACTORY EPITHELIUM;
RX   MEDLINE=20442421; PubMed=10984544;
RA   Meyer M.R., Angele A., Kremmer E., Kaupp U.B., Mueller F.;
RT   "A cGMP-signalling pathway in a subset of olfactory sensory neurons.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:10595-10600(2000).
DR   EMBL; AJ272428; CAC09430.1; -.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR001622; Channel_pore_K.
DR   InterPro; IPR000595; cNMP_binding.
DR   Pfam; PF00027; cNMP_binding; 1.
DR   Pfam; PF00520; ion_trans; 1.
DR   SMART; SM00100; cNMP; 1.
DR   PROSITE; PS00888; CNMP_BINDING_1; 1.
DR   PROSITE; PS50042; CNMP_BINDING_3; 1.
SQ   SEQUENCE   670 AA;  76420 MW;  65203ADE66F85F62 CRC64;


  Query Match             52.8%;  Score 1578;  DB 11;  Length 670;
  Best Local Similarity   54.5%;  Pred. No. 7e-110;
```

```
  Matches  300;  Conservative  105;  Mismatches  125;  Indels   20;  Gaps    5;

Qy      14 PPAPSKARKLLP-VLDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHGYLVAWLVL 72
           ||    | ||  | |:||| : || ||  :  || ||  :|||||||  :||  :|  ||||
Db     127 PPEGGKGRKKDPIVVDPSSNIYYRWLTAIALPVFYNWCLLVCRACFDELQSEHLTLWLVL 186

Qy      73 DYTSDLLYLLDMVVRFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLASLMPTDVVYVRL 132
           ||::| ||::||:||  |||||||::| |  |:   | :|  | ||: ||:|||: |:::|
Db     187 DYSADALYVVDMLVRARTGFLEQGLMVRDTKRLWKHYTKTLHFKLDILSLIPTDLAYLKL 246

Qy     133 GPHTPTLRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSCLYFALSR 192
           | : | || || ||  |||| |||||||| ||| |||  |:||  ::||||:|:|||:|:
Db     247 GMNYPELRFNRLLRFSRLFEFFDRTETRTNYPNVFRIGNLVLYTLIIIHWNACIYFAISK 306

Qy     193 YLGFGRDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAREEEYLFMVGDFL 252
           ::||| |:|||: ::| : || |:|:|| |:||| |||:|:|||| ::|||||:| |||
Db     307 FIGFGTDSWVYPNTSKPEYGRLSRKYIYSLYWSTLTLTTIGETPPPVKDEEYLFVVIDFL 366

Qy     253 LAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVIDWYQHLQI 312
           : |: ||||:|:: |:| ||| : | |      :|:||: : | : || ||| |: :|
Db     367 VGVLIFATIVGNVGSMISNMNASRAEFQAKIDSIKQYMQFRKVTKDLETRVIRWFDYLWA 426

Qy     313 NKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQPQTYSPG 372
           |:|  :|  :|::||::|:||:|::||| || :|:|||:||| || ||||||:|  :|||
Db     427 NRKTVDEKEVLKNLPDKLKAEIAINVHLDTLKKVRIFQDCEAGLLVELVLKLRPAVFSPG 486

Qy     373 EYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIINIKGNMSGNRRTA 432
           :|:|:|||||:|||||:||:|||||||:||: ||  | |||||||:||||: ||||||||
Db     487 DYICKKGDIGREMYIIKEGKLAVVADDGVTQFVVLSDGSYFGEISILNIKGSKSGNRRTA 546

Qy     433 NIKSLGYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLD-----VNAEAAEI 487
           ||:|:||||||||||:|| | |:||| |:  :|||||:||:| | :|       |:|  |
Db     547 NIRSIGYSDLFCLSKDDLMETLTEYPDAKRALEEKGRQILMKDNLIDDDLVTARADARNI 606

Qy     488 ALQEATESRLRGLDQQLDDLQTKFARLLAELESSALKIAYRIERLEWQ-TR-------EW 539
                 | ::  |:  || |||:|||||||  :| :|:  |: :|| | ||       :
Db     607 ------EEKVEYLESSLDGLQTRFARLLAEYSASQMKLKQRLSQLESQMTRRGHGFSPDR 660

Qy     540 PMPEDLAEAD 549
             || ::||
Db     661 ENSEDASKAD 670


RESULT   4
Q9N0H4
ID   Q9N0H4      PRELIMINARY;      PRT;   686 AA.
AC   Q9N0H4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CGMP-GATED CATION CHANNEL.
GN   CNG-1.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ratcliffe C.F., Brammar W.J., Conley E.C.;
RT   "Cloning of cDNAs Encoding cGMP-Gated Cation Channels from Intact
RT   Coronary Arteries and Human Vascular Endothelial Cells.";
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U85404; AAD05039.1; -.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR001622; Channel_pore_K.
DR   InterPro; IPR000595; cNMP_binding.
DR   Pfam; PF00027; cNMP_binding; 1.
DR   Pfam; PF00520; ion_trans; 1.
DR   SMART; SM00100; cNMP; 1.
DR   PROSITE; PS00888; CNMP_BINDING_1; 1.
DR   PROSITE; PS00889; CNMP_BINDING_2; 1.
DR   PROSITE; PS50042; CNMP_BINDING_3; 1.
SQ   SEQUENCE   686 AA;  79125 MW;  DB86A7C16CE79D3F CRC64;


  Query Match             51.5%;  Score 1538.5;  DB 6;  Length 686;
  Best Local Similarity   54.0%;  Pred. No. 6.7e-107;
  Matches  285;  Conservative  113;  Mismatches  129;  Indels    1;  Gaps    1;

Qy       6 KVKTTESSPPAPSKARKLLPVLDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHGY 65
           | |  | |     : :| : |:||||: || ||  :  |||||  ::: |||| :||  |
Db     133 KKKKEEKSKDKKEEEKKEVVVIDPSGNTYYNWLFCITLPVMYNWTMVIARACFDELQSDY 192

Qy      66 LVAWLVLDYTSDLLYLLDMVVRFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLASLMPT 125
           |  ||:||| ||::||:|| ||  ||:||||:|| :: :: ::|     | ||: ||:||
Db     193 LEYWLILDYVSDIVYLIDMFVRTRTGYLEQGLLVKEELKLINKYKSNLQFKLDVLSLIPT 252

Qy     126 DVVYVRLGPHTPTLRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSC 185
           |::| :|| : | :|||| ||  |:|| | |||||| ||| |||: |:::|| ::||||:|
Db     253 DLLYFKLGWNYPEIRLNRLLRFSRMFEFFQRTETRTNYPNIFRISNLVMYIVIIIHWNAC 312

Qy     186 LYFALSRYLGFGRDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAREEEYL 245
           :::::|: :||| | ||||    | | || |:|:|| |:||| |||:|:|||| |: ||:
Db     313 VFYSISKAIGFGNDTWVYPDINDPEFGRLARKYVYSLYWSTLTLTTIGETPPPVRDSEYV 372

Qy     246 FMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVID 305
           |:| |||: |: ||||:|:: |:| ||| | | |       :|:||  ::|:: :|:|||
Db     373 FVVVDFLIGVLIFATIVGNIGSMISNMNAARAEFQARIDAIKQYMHFRNVSKDMEKRVIK 432

Qy     306 WYQHLQINKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQ 365
           |: :|  |||  :|  :|::||::||||:|::||| || :|:|| :||| || |||||||
Db     433 WFDYLWTNKKTVDEKEVLKYLPDKLRAEIAINVHLDTLKKVRIFADCEAGLLVELVLKLQ 492

Qy     366 PQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIINIKGNM 425
           || ||||:|:|:|||||:|||||:||:|||||||:||: |||||||:|||:
Db     493 PQVYSPGDYICKKGDIGREMYIIKEGKLAVVADDGVTQFVVLSDGSYFGEISILNIKGSK 552

Qy     426 SGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLDVNAEAA 485
           :|||||||||:||||||||||:|| | |:||| |:|::||||::||:|   ||:|   |
Db     553 AGNRRTANIKSIGYSDLFCLSKDDLMEALTEYPDAKTMLEEKGKQILMKDGLLDLNIANA 612

Qy     486 EIALQEATESRLRGLDQQLDDLQTKFARLLAELESSALKIAYRIERLE 533
             ::  | ::  ::  :| |||:|||:||| ||   |:  |: ::|
Db     613 GSDPKD-LEEKVTRMEGSVDLLQTRFARILAEYESMQQKLKQRLTKVE 659


RESULT   5
P90800
ID   P90800      PRELIMINARY;      PRT;   733 AA.
AC   P90800;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CYCLIC NUCLEOTIDE-GATED CHANNEL.
GN   TAX-4.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97048186; PubMed=8893027;
RA   Komatsu H., Mori I., Rhee J.S., Akaike N., Ohshima Y.;
RT   "Mutations in cyclic nucleotide-gated channel lead to abnormal
RT   thermosensation and chemosensation in C. elegans.";
RL   Neuron 17:707-718(1996).
DR   EMBL; D86922; BAA13180.1; -.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR001622; Channel_pore_K.
DR   InterPro; IPR000595; cNMP_binding.
DR   Pfam; PF00027; cNMP_binding; 1.
DR   Pfam; PF00520; ion_trans; 1.
DR   SMART; SM00100; cNMP; 1.
DR   PROSITE; PS00888; CNMP_BINDING_1; 1.
DR   PROSITE; PS00889; CNMP_BINDING_2; 1.
DR   PROSITE; PS50042; CNMP_BINDING_3; 1.
```

```
SQ   SEQUENCE   733 AA;  83892 MW;  1EC7EB0179F9355E CRC64;

  Query Match             37.5%;  Score 1119.5;  DB 5;  Length 733;
  Best Local Similarity   39.9%;  Pred. No. 2e-75;
  Matches  236;  Conservative  112;  Mismatches  165;  Indels   79;  Gaps    9;

Qy      11 ESSPPAPSKARKLLPVL-----DPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDL---- 61
Db      93 EVPPPKESYAVRIRKYLANYTQDPSTDNFYYWTCVVTVAYIYNLLFVIARQVFNDLIGPS 152

Qy      62 -------QHGYL-----------------------------------VAWLVLDYTSD 77
Db     153 SQSLCRFYNGTLNSTTQVECTYNMLTNMKEMPTYSQYPDLGWSKYWHFRMLWVFFDLLMD 212

Qy      78 LLYLLDMVVRFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLASLMPTDVVYVRLG---P 134
Db     213 CVYLIDTFLNYRMGYMDQGLVVREAEKVTKAYWQSKQYRIDGISLIPLDYI---LGWPIP 269

Qy     135 H-----TPTLRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSCLYFA 189
Db     270 YINWRGLPILRLNRLIRYKRVRNCLERTETRSSMPNAFRVVVVVWYIVIIIHWNACLYFW 329

Qy     190 LSRYLGFGRDAWVY--------PDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPARE 241
Db     330 ISEWIGLGTDAWVYGHLNKQSLPDDIT---DTLLRRYVYSFYWSTLILTTIGEVPSPVRN 386

Qy     242 EEYLFMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLER 301
Db     387 IEYAFVTLDLMCGVLIFATIVGNVGSMISNMSAARTEFQNKMDGIKQYMELRKVSKQLEI 446

Qy     302 RVIDWYQHLQINKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELV 361
Db     447 RVIKWFDYLWTNKQSLSDQQVLKVLPDKLQAEIAMQVHFETLRKVRIFQDCEAGLLAELV 506

Qy     362 LKLQPQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIINI 421
Db     507 LKLQLQVFSPGDFICKKGDIGREMYIVKRGRLQVVDDDGKKVFVTLQEGSVFGELSILNI 566

Qy     422 KGNMSGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLDVN 481
Db     567 AGSKNGNRRTANVRSVGYTDLFVLSKTDLWNALREYPDARKLLLAKGREILKKDNLLDEN 626

Qy     482 AEAAEIALQEATESRLRGLDQQLDDLQTKFARLLAELESSALKIAYRIERLE 533
Db     627 APEEQKTVEEIAEH----LNNAVKVLQTRMARLIVEHSSTEGKLMKRIEMLE 674


RESULT   6
O97119
ID   O97119      PRELIMINARY;      PRT;   905 AA.
AC   O97119;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CYCLIC NUCLEOTIDE-GATED ION CHANNEL LCNG1.
GN   LCNG1.
OS   Limulus polyphemus (Atlantic horseshoe crab).
OC   Eukaryota; Metazoa; Arthropoda; Chelicerata; Merostomata; Xiphosura;
OC   Limulidae; Limulus.
OX   NCBI_TaxID=6850;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN AND VENTRAL EYE NERVE;
RX   MEDLINE=99127907; PubMed=9930717;
RA   Chen F.H., Ukhanova M., Thomas D., Afshar G., Tanda S., Battelle B.A.,
RA   Payne R.;
RT   "Molecular cloning of a putative cyclic nucleotide-gated ion channel
RT   cDNA from Limulus polyphemus.";
RL   J. Neurochem. 72:461-471(1999).
DR   EMBL; AF091302; AAD16099.1; -.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR001622; Channel_pore_K.
DR   InterPro; IPR000595; cNMP_binding.
DR   Pfam; PF00027; cNMP_binding; 1.
DR   Pfam; PF00520; ion_trans; 1.
DR   SMART; SM00100; cNMP; 1.
DR   PROSITE; PS00888; CNMP_BINDING_1; 1.
DR   PROSITE; PS00889; CNMP_BINDING_2; UNKNOWN_1.
DR   PROSITE; PS50042; CNMP_BINDING_3; 1.
SQ   SEQUENCE   905 AA;  102956 MW;  C9D4C13FC5F5412D CRC64;

  Query Match             33.1%;  Score 990.5;  DB 5;  Length 905;
  Best Local Similarity   37.7%;  Pred. No. 1.3e-65;
  Matches  202;  Conservative  111;  Mismatches  176;  Indels   47;  Gaps    6;

Qy      26 VLDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHGYLVAWLVLDYTSDLLYLLDMV 85
Db     191 VFDPSGSLSYYWSMIVSVAFLYNFWVVIFRFAFNEIKANTIVAWLTLDYFADILYALDIV 250

Qy      86 VRFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLASLMPTDVVYVRLGPHTPTLRLNRFL 145
Db     251 FHFRTGYLEEGVLQTDTIKLRHHYMNSTLFYIDCLCLLPIDFLYLSIG-FNSILRCLRLV 309

Qy     146 RAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSCLYFALSRYLGFGRDAWVYPD 205
Db     310 KIYRFWTFLDRTERHTNYPNVFRTVILTHYVLVIFHWNACLFHIISVNGGFGTRSWF--N 367

Qy     206 PAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAREEEYLFMVGDFLLAVMGFATIMGSM 265
Db     368 PRDETCNDVVCEYLRSFYWSTLALTTIGDLPSPRTKGEYLFLVVELVFGLFLFAAVLGHV 427

Qy     266 SSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVIDWYQHLQINKKMTNEVAILQH 325
Db     428 ANIVTNVSTARKEFQARLDVVKTYMRMRRVPDHLQNKVIKWFDYLWVTQKSSDEDRSVGF 487

Qy     326 LPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQPQTYSPGEYVCRKGDIGQEM 385
Db     488 LPDKLKAEIAIHVHLNTLKQVEIFQNTEAGFLCELVLRLRPVLFSPGDYICRKGEVGKEM 547

Qy     386 YIIREGQLAVVADDGITQYAVLGAGLYFGEISIINIKGNMSGNRRTANIKSLGYSDLFCL 445
Db     548 YIVNRGRLEVVTDNGKTVLATLRAGSYFGEISILNM--STTGNRRTASVRSVGYSDLFCL 605

Qy     446 SKEDLREVLSEYPQAQT------------------------------IMEEKGREILLK 474
Db     606 NKQDMWDVLKDYPAARDRLESIAVKRLEKYRKDPLQNIAMSRSHSTPGLLESHGQVPLEN 665

Qy     475 M----NKLDVNAEAAEIALQE-------ATESRLRGLDQQLDDLQTKFARLLAELE 519
Db     666 MPLYPREPSTTASAATLLSDDECRQNVRATASGTRSRRTSVDDIQQRGTPLYEQIE 721


RESULT   7
Q9W202
ID   Q9W202      PRELIMINARY;      PRT;   702 AA.
AC   Q9W202;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   CG3536 PROTEIN.
GN   CG3536.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
```

```
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003458; AAF46898.1; -.
DR   FlyBase; FBgn0034775; CG3536.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR001622; Channel_pore_K.
DR   InterPro; IPR000595; cNMP_binding.
DR   Pfam; PF00027; cNMP_binding; 1.
DR   Pfam; PF00520; ion_trans; 1.
DR   SMART; SM00100; cNMP; 1.
DR   PROSITE; PS00888; CNMP_BINDING_1; 2.
DR   PROSITE; PS00889; CNMP_BINDING_2; 2.
DR   PROSITE; PS50042; CNMP_BINDING_3; 1.
SQ   SEQUENCE   702 AA;  80289 MW;  BF175E4CD95D86CF CRC64;
```

```
  Query Match             31.7%;  Score 947;  DB 5;  Length 702;
  Best Local Similarity   42.2%;  Pred. No. 1.6e-62;
  Matches  185;  Conservative   95;  Mismatches  150;  Indels    8;  Gaps    4;

Qy      40 TMVFPVMYNLIILVCRACFPDLQHGYLVAWLVLDYTSDLLYLLDMVVRFHTGFLEQGILV 99
           :| |  :||  ::: |  | ::    :  |  ||| || |||:|::  | ||:|| |:|
Db       4 SMAF--LYNFWVIIYRFAFQEINRRTIAIWFCLDYLSDFLYLIDILFHFRTGYLEDGVLQ 61

Qy     100 VDKGRISSRYVRTWSFFLDLASLMPTDVVYVRLGPHTPTLRLNRFLRAPRLFEAFDRTET 159
             |  :: : |: :  |::|   |:| | :|: :|     ||  | ::  | :   ||||
Db      62 TDALKLRTHYMNSTIFYIDCLCLLPLDFLYLSIG-FNSILRSFRLVKIYRFWAFMDRTER 120

Qy     160 RTAYPNAFRIAKLMLYIFVVIHWNSCLYFALSRYLGFGRDAWVYPDPAQPGFERLRRQYL 219
           | ||| ||   |: |: |: ||| |||  : :  |||   ||| |         : :|||
Db     121 HTNYPNLFRSTALIHYLLVIFHWNGCLYHIIHKNNGFGSRNWVYHDSESAD---VVKQYL 177

Qy     220 YSFYFSTLILTTVGDTPPPAREEEYLFMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAF 279
           |:|: || |||:|| | |  : ||:|::   |  :| |||::| ::::: ::: |   |
Db     178 QSYYWCTLALTTIGDLPKPRSKGEYVFVILQLLFGLMLFATVLGHVANIVTSVSAARKEF 237

Qy     280 YPDHALVKKYMKLQHVNRKLERRVIDWYQHLQINKKMTNEVAILQHLPERLRAEVAVSVH 339
             || ||::: |   |: :|| |: :| : :| ::|   :  ||::|:||:|::||
Db     238 QAKLDGVKTYMRMRRVPNHLQVKVIKWFDYLWLTQKCSDEERAVSCLPDKLKAEIAINVH 297

Qy     340 LSTLSRVQIFQNCEASLLEELVLKLQPQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADD 399
```

```
           | || ||:|||| ||  | ||||:|:|  :|||:|:||||::|:||||:  |:| ||||:
Db     298 LDTLKRVEIFQNTEAGFLCELVLRLRPVLFSPGDYICRKGEVGKEMYIVNRGRLQVVADN 357

Qy     400 GITQYAVLGAGLYFGEISIINIKGNMSGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQ 459
           | |  | | || |||||||:|:    :|||||:::|:|:|||||| |||:|: :|| |||
Db     358 GKTVMASLKAGSYFGEISILNM--GTAGNRRTASVRSVGYSDLFVLSKKDMWDVLKEYPA 415

Qy     460 AQTIMEEKGREILLKMNK 477
           |:  :|    : | |  |
Db     416 ARVRLESIAVKRLEKYKK 433


RESULT   8
Q9U5E2
ID   Q9U5E2      PRELIMINARY;      PRT;  1463 AA.
AC   Q9U5E2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   CNG CHANNEL-LIKE.
GN   CNGL OR CG9176.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Miyazu M., Tanimura T.;
RT   "A putative CNG-channel of Drosophila melanogaster.";
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB030577; BAA89278.1; -.
DR   FlyBase; FBgn0029090; cngl.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR001622; Channel_pore_K.
DR   InterPro; IPR000595; cNMP_binding.
DR   Pfam; PF00027; cNMP_binding; 1.
DR   Pfam; PF00520; ion_trans; 1.
DR   SMART; SM00100; cNMP; 1.
DR   PROSITE; PS00888; CNMP_BINDING_1; UNKNOWN_1.
DR   PROSITE; PS00889; CNMP_BINDING_2; UNKNOWN_1.
DR   PROSITE; PS50042; CNMP_BINDING_3; 1.
SQ   SEQUENCE   1463 AA;  159935 MW;  D902602E351D42C2 CRC64;
```

```
  Query Match             29.4%;  Score 879;  DB 5;  Length 1463;
  Best Local Similarity   33.0%;  Pred. No. 6.1e-57;
  Matches  202;  Conservative  123;  Mismatches  219;  Indels   68;  Gaps   12;

Qy      21 RKLLPVLDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHGYLVAWLVLDYTSDLLY 80
           :|   |::|  ::|::||  :   |:|||  |: |  ||:||      ||: |  :|:::
Db     112 QKRRSVVNPDENFYFYWLMMLTVCVLYNLWTLIVRQSFPELQQSVPTFWLICDSMTDVVF 171

Qy      81 LLDMVVRFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLASLMPTDVVYVRLGPHTPTLR 140
           :||::|:  ||:||||::| |  :::  || :  |  |: :|:| |:: :::| | | ||
Db     172 ILDIIVQLRTGYLEQGLMVYDDRKLACHYVHSRDFIFDMIALIPLDLLQLKMGTH-PLLR 230

Qy     141 LNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSCLYFALSRYLGFGRDA 200
             || :  |    :   |:|| :|| :|:  |:  : :: || , | || ||   ||  |
Db     231 FTRFFKVYRSVRFYYIVESRTVWPNLWRVVNLIHILLILAHWFGCFYFLLSEAEGFQGD- 289

Qy     201 WVYPDPAQPG-FERLRRQYLYSFYFSTLILTTVGDTPPPAREEEYLFMVGDFLLAVMGFA 259
           |||  | :|| :  | |:|| | |:||| |||:|| | |    ||:| :  :|: |  ||
Db     290 WVY--PYRPGDYPTLTRKYLGSLYWSTLTLTTIGDLPTPETNAEYIFTIVSYLIGVFIFA 347

Qy     260 TIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVIDWYQHLQINKKMTNE 319
           ||:| : :|| | |      |       | ||:   |   ::|||: || :     ::
Db     348 TIVGQVGNVITNRNANRLEFERLLDGAKTYMRHHKVPGGMKRRVLRWYDYSWSRGRIQGG 407

Qy     320 VAI---LQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQPQTYSPGEYVC 376
           |   |  ||:::|: |:|: |:|| | :| ||| |:   | :||||::   ::||: :|
Db     408 GDINTALGLLPDKLKTELALHVNLSVLKKVTIFQECQPEFLHDLVLKMKAYIFTPGDSIC 467
```

```
Qy     377 RKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIINIKGNMSGNRRTANIKS 436
           |||:: :||:|| :| | |::: |      : || :|||| |:|: |    |:|||:::|
Db     468 RKGEVAREMFIIADGILEVLSETG-KVLTTMKAGDFFGEIGILNLDGL---NKRTADVRS 523

Qy     437 LGYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLK---MNKLDVNAEAAEIALQEAT 493
           :|||:|| ||:||:   : :|| || |::  ||: |::   :||    |:: : |   |
Db     524 VGYSELFSLSREDVLAAMKDYPDAQEILQTLGRKRLMEVRCVNKKYAKAQSDKEAAAYAA 583

Qy     494 ------ESRLRGLDQQLDDLQTKFARLLAELESSALKIAYRI-----------ERLEWQT 536
                :|  :|   | |  :   :: : :     :|    :           | || |
Db     584 AHPHHHQSHHQGQVHQSDSSENSASKKIVDKLKHDVKGFRNVLKKSRTSRKSDESLEMQP 643

Qy     537 REWPMPE------------------DLAEADDEGEPEEGT----------------SK 560
           |                    | ||| ||   :  :                    :
Db     644 LHNTSPRGSKILLKRMSRVRSDEKDADSAEAKDELHDKTPSPIGAGLPLLQRLRLLKEKQ 703

Qy     561 DEEGRASQEGPP 572
           | | || :  ||
Db     704 DREERAVKSTPP 715


RESULT   9
Q9VXV8
ID   Q9VXV8      PRELIMINARY;      PRT;  1693 AA.
AC   Q9VXV8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   CG9176 PROTEIN.
GN   CNGL OR CG9176.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003498; AAF48450.1; -.
DR   FlyBase; FBgn0029090; cngl.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR001622; Channel_pore_K.
DR   InterPro; IPR000595; cNMP_binding.
DR   Pfam; PF00027; cNMP_binding; 1.
DR   Pfam; PF00520; ion_trans; 1.
DR   SMART; SM00100; cNMP; 1.
DR   PROSITE; PS00888; CNMP_BINDING_1; 1.
DR   PROSITE; PS00889; CNMP_BINDING_2; 1.
DR   PROSITE; PS50042; CNMP_BINDING_3; 1.
SQ   SEQUENCE   1693 AA;  182854 MW;  697857836975C4D4 CRC64;


  Query Match             27.8%;  Score 829.5;  DB 5;  Length 1693;
  Best Local Similarity   33.3%;  Pred. No. 3.9e-53;
  Matches  197;  Conservative  116;  Mismatches  204;  Indels   75;  Gaps   14;

Qy      45 VMYNLIILVCRACFPDLQHGYLVAWLVLDYTSDLLYLLDMVVRFHTGFLEQGILVVDKGR 104
           |:|||  |: |  ||:||      ||: |  :|::::||::|:  ||:||||::| |  :
Db       7 VLYNLWTLIVRQSFPELQQSVPTFWLICDSMTDVVFILDIIVQLRTGYLEQGLMVYDDRK 66

Qy     105 ISSRYVRTWSFFLDLASLMPTDVVYVRLGPHTPTLRLNRFLRAPRLFEAFDRTETRTAYP 164
           ::  || :  |  |: :|:| |:: :::| | | ||  |  |: | :      |:|| :|
Db      67 LACHYVHSRDFIFDMIALIPLDLLQLKMGTH-PLLRFTRVYRSVRFYYI---VESRTVWP 122

Qy     165 NAFRIAKLMLYIFVVIHWNSCLYFALSRYLGFGRDAWVYPDPAQPG-FERLRRQYLYSFY 223
           | :|:  |:  : :: ||  | || ||   ||  | |||  | :|| :  | |:|| | |
Db     123 NLWRVVNLIHILLILAHWFGCFYFLLSEAEGFQGD-WVY--PYRPGDYATLTRKYLGSLY 179

Qy     224 FSTLILTTVGDTPPPAREEEYLFMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDH 283
           :||| |||:|| | |    ||:| :  :|: |  ||||:| : :|| | |     |
Db     180 WSTLTLTTIGDLPTPETNAEYIFTIVSYLIGVFIFATIVGQVGNVITNRNANRLEFERLL 239

Qy     284 ALVKKYMKLQHVNRKLERRVIDWYQHLQINKKMTNEVAI---LQHLPERLRAEVAVSVHL 340
           | ||:   |   ::|||: || :    .::     |   |  ||::|: |:|: |:|
Db     240 DGAKTYMRHHKVPGGMKRRVLRWYDYSWSRGRIQGGGDINTALGLLPDKLKTELALHVNL 299

Qy     341 STLSRVQIFQNCEASLLEELVLKLQPQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADDG 400
           | | :| ||| |:   | :||||::   ::||: :||||:: :|||:|| :| | |::: |
Db     300 SVLKKVTIFQECQPEFLHDLVLKMKAYIFTPGDSICRKGEVAREMFIIADGILEVLSETG 359

Qy     401 ITQYAVLGAGLYFGEISIINIKGNMSGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQA 460
                 : || :|||| |:|: |    |:|||:::|:|||:|| ||:||:   : :|| |
Db     360 -KVLTTMKAGDFFGEIGILNLDGL---NKRTADVRSVGYSELFSLSREDVLAAMKDYPDA 415

Qy     461 QTIMEEKGREILLK---MNKLDVNAEAAEIALQEAT------ESRLRGLDQQLDDLQTKF 511
           | |::  ||: |::   :||    |:: : |   |       :|  :|   | |  :
Db     416 QEILQTLGRKRLMEVRCVNKKYAKAQSDKEAAAYAAAHPHHHQSHHQGQVHQSDSSENSA 475

Qy     512 ARLLAELESSALKIAYRI-----------ERLEWQTREWPMPE---------------- 543
           :: : :     :|    :           | || |      |
Db     476 SKKIVDKLKHDVKGFRNVLKKSRTSRKSDESLEMQPLHNTSPRGSKILLKRMSRVRSDEK 535

Qy     544 --DLAEADDE----------GEP----------EEGTSKDEEGRASQEGPP 572
             | ||| ||          | |          :: : ::|  |   || |
Db     536 DADSAEAKDELHDKTPSPIGAGLPLLQRLRLLKEKQESIQEEPEREFSEGFP 587


RESULT  10
Q9QX26
ID   Q9QX26      PRELIMINARY;      PRT;   252 AA.
AC   Q9QX26;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   CYCLIC NUCLEOTIDE-GATED CATION CHANNEL (FRAGMENT).
GN   CNG3.
OS   Rattus norvegicus (Rat).
```

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Qiu W., Guggino S.E.;
RT   "Cyclic nucleotide-gated cation channel (CNG3) expression in rat
RT   kidney.";
RL   Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF031943; AAB87065.1; -.
DR   InterPro; IPR001622; Channel_pore_K.
DR   InterPro; IPR000595; cNMP_binding.
DR   Pfam; PF00027; cNMP_binding; 1.
DR   SMART; SM00100; cNMP; 1.
DR   PROSITE; PS00888; CNMP_BINDING_1; 1.
DR   PROSITE; PS00889; CNMP_BINDING_2; 1.
DR   PROSITE; PS50042; CNMP_BINDING_3; 1.
FT   NON_TER       1      1
FT   NON_TER     252    252
SQ   SEQUENCE   252 AA;  28677 MW;  C59A1CF7654AA884 CRC64;


  Query Match             27.7%;  Score 829;  DB 11;  Length 252;
  Best Local Similarity   59.9%;  Pred. No. 2.6e-54;
  Matches  151;  Conservative   55;  Mismatches   46;  Indels    0;  Gaps    0;

Qy     208 QPGFERLRRQYLYSFYFSTLILTTVGDTPPPAREEEYLFMVGDFLLAVMGFATIMGSMSS 267
           :| : || |:|:|| |:||| |||:|:|||| ::|||||:| |||: |: ||||:|:: |
Db       1 KPEYGRLSRKYIYSLYWSTLTLTTIGETPPPVKDEEYLFVVIDFLVGVLIFATIVGNVGS 60

Qy     268 VIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVIDWYQHLQINKKMTNEVAILQHLP 327
           :| ||| : | |      :|:||: : | : || ||| |: :|  |:|  :|  :|::||
Db      61 MISNMNASRAEFQAKIDSIKQYMQFRKVTKDLETRVIRWFDYLWANRKTVDEKEVLKNLP 120

Qy     328 ERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQPQTYSPGEYVCRKGDIGQEMYI 387
           ::|:||:|::||| || :|:|||:||| || |||||:|  :|||:|:|:|||||:||||
Db     121 DKLKAEIAINVHLDTLKKVRIFQDCEAGLLVELVLKLRPAVFSPGDYICKKGDIGREMYI 180

Qy     388 IREGQLAVVADDGITQYAVLGAGLYFGEISIINIKGNMSGNRRTANIKSLGYSDLFCLSK 447
           |:||:|||||||:||: ||  | |||||||:|||: |||||||||:|:|||||||||
Db     181 IKEGKLAVVADDGVTQFVVLSDGSYFGEISILNIKGSKSGNRRTANIRSIGYSDLFCLSK 240

Qy     448 EDLREVLSEYPQ 459
           :|| | |:|||:
Db     241 DDLMETLTEYPE 252


RESULT  11
P90975
ID   P90975      PRELIMINARY;      PRT;   800 AA.
AC   P90975;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   CYCLIC NUCLEOTIDE-GATED CHANNEL PROTEIN TAX-2.
GN   TAX-2 OR F36F2.5.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97048185; PubMed=8893026;
RA   Coburn C.M., Bargmann C.I.;
RT   "A putative cyclic nucleotide-gated channel is required for sensory
RT   development and function in C. elegans.";
RL   Neuron 17:695-706(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Kershaw J.;
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U73476; AAB41492.1; -.
DR   EMBL; Z93778; CAB07847.1; -.
DR   EMBL; Z81532; CAB07847.1; JOINED.
DR   HSSP; P00514; 1BPK.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR001622; Channel_pore_K.
DR   InterPro; IPR000595; cNMP_binding.
DR   InterPro; IPR001202; WW.
DR   Pfam; PF00027; cNMP_binding; 1.
DR   Pfam; PF00520; ion_trans; 1.
DR   SMART; SM00100; cNMP; 1.
DR   PROSITE; PS00888; CNMP_BINDING_1; 1.
DR   PROSITE; PS00889; CNMP_BINDING_2; UNKNOWN_1.
DR   PROSITE; PS50042; CNMP_BINDING_3; 1.
DR   PROSITE; PS01159; WW_DOMAIN_1; UNKNOWN_1.
SQ   SEQUENCE   800 AA;  92462 MW;  D0498AD2C340FDAA CRC64;


  Query Match             26.4%;  Score 788.5;  DB 5;  Length 800;
  Best Local Similarity   32.4%;  Pred. No. 1.5e-50;
  Matches  184;  Conservative  109;  Mismatches  228;  Indels   47;  Gaps   10;

Qy      29 PSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHGYLVAWLVLDYTSDLLYLLDM-VVR 87
           | | :|  ||: :    ::|   :  |: :|       : | ::||: ||:|::|| :::
Db     216 PYGRFYMTWLSLVTLCFLFNAFCIPLRSSYPYQTADNWMYWFIVDYSCDLVYVIDMLLIK 275

Qy      88 FHTGFLEQGILVVDKGRISSRYVRTWSFFLDLASLMPTDVVYVRLGPHTPTLRLNRFLRA 147
             |   || |          |: | :| ||: |::|||::|   |   |  |:|| |:
Db     276 PRLRFTRGGIQVKIYKDTQRHYLMTRTFKLDILSILPTDLMYFFFG-KMPIWRINRVLKI 334

Qy     148 PRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSCLYFALSRYLGFGRDA------- 200
             :  ||  :   | | | |||: : |:  :|| |||:|: ||    ||: |
Db     335 NSFWLLFDMLDNSFANPYAIRIARTLSYMIYIIHCNSCVYYKLSALQAFGQIAYLENGKW 394

Qy     201 ----WVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAREEEYLFMVGDFLLAVM 256
               ||| :          |:  |||:  : |: |: | |    ||::|   ::: |
Db     395 YLNKWVYNNQG--------NAYIRCFYFTAAVATSTGNNPAPTNVIEYIYMTCSWMMGVF 446

Qy     257 GFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVIDWYQHLQINKKM 316
           || ::| :  :: | |      |     |      |   :  :   || ||: :    :|
Db     447 VFALLLGQIRDIVSNANRNREEFQRKMDLALGECKKLGLKMETTNRVRDWFIYTWQQQKT 506

Qy     317 TNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQPQTYSPGEYVC 376
           :|  ::: || :|: ::|:|:|| :|||:||:||:|: :|| :|||||:|  : ||: :|
Db     507 LDEKKLIEKLPLKLQTDLALSVHYTTLSKVQLFQDCDRALLRDLVLKLRPVIFLPGDMIC 566

Qy     377 RKGDIGQEMYIIREGQLAVVADDGITQ-YAVLGAGLYFGEISIINIKGNMSGNRRTANIK 435
           |||:|:|||||| :| | ||  |   : :| |  | |||||:: | ||   |||||:|:
Db     567 LKGDVGKEMYIINQGILQVVGGDHNEKIFAELAQGAVFGEISLLAIGGN---NRRTASIR 623

Qy     436 SLGYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLDVNAEAAEIALQEATES 495
           : ||  || |:|||| :|: |||||||:  |   :|    | |   | :|  |   |
Db     624 AKGYCTLFVLAKEDLNDVIRYYPQAQTILRRKAAAMLKNDKKSDEKTE--KIKAQAELED 681

Qy     496 RLRGLDQQLDDLQTKFARLLAELESSALKIAYRI--ERLE----------WQT------- 536
           | :   :|:  | |  | :    |:  ::   ::  |  |           | |
Db     682 RCKINPRQVPKLITLIANMTEMNENKGVQELKKVIEEETEKSRRQSIYYPWSTLQRDDDD 741

Qy     537 -REWPMPEDLAEADDEGEPEEGTSKDEE 563
            ||   |||::  :: : :     |:|
Db     742 EEEWNDEEDLSDVGEDFDLDPTNHSDDE 769


RESULT  12
O62237
ID   O62237      PRELIMINARY;      PRT;   800 AA.
AC   O62237;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   F36F2.5 PROTEIN.
GN   F36F2.5.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
```

```
CX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Cottage A.;
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; Z81532; CAB04328.1; -.
DR   EMBL; Z93778; CAB04328.1; JOINED.
DR   HSSP; P00514; 1BPK.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR001622; Channel_pore_K.
DR   InterPro; IPR000595; cNMP_binding.
DR   InterPro; IPR001202; WW.
DR   Pfam; PF00027; cNMP_binding; 1.
DR   Pfam; PF00520; ion_trans; 1.
DR   SMART; SM00100; cNMP; 1.
DR   PROSITE; PS00888; CNMP_BINDING_1; 1.
DR   PROSITE; PS00889; CNMP_BINDING_2; UNKNOWN_1.
DR   PROSITE; PS50042; CNMP_BINDING_3; 1.
DR   PROSITE; PS01159; WW_DOMAIN_1; UNKNOWN_1.
SQ   SEQUENCE   800 AA;  92534 MW;  8A49B44F8940977F CRC64;
```

```
  Query Match           26.4%;  Score 788.5;  DB 5;  Length 800;
  Best Local Similarity  32.4%;  Pred. No. 1.5e-50;
  Matches  184;  Conservative  109;  Mismatches  228;  Indels   47;  Gaps   10;

Qy      29 PSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHGYLVAWLVLDYTSDLLYLLDM-VVR 87
           | | :|  ||: :    ::|   :  |: :|     : | ::||: ||:|::|| ::: 
Db     216 PYGRFYMTWLSLVTLCFLFNAFCIPLRSSYPYQTADNWMYWFIVDYSCDLVYVIDMLLIK 275

Qy      88 FHTGFLEQGILVVDKGRISSRYVRTWSFFLDLASLMPTDVVYVRLGPHTPTLRLNRFLRA 147
             |   || |         |: | :| ||: |::|||::|   |   |  |:|| |: 
Db     276 PRLRFTRGGIQVKIYKDTQRHYLMTRTFKLDILSILPTDLMYFFFG-KMPIWRINRVLKI 334

Qy     148 PRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSCLYFALSRYLGFGRDA------- 200
             :  ||  :   | | | |||: : |:  :|| |||:|: ||    ||: |
Db     335 NSFWLLFDMLDNSFANPYAIRIARTLSYMIYIIHCNSCVYYKLSALQAFGQIAYLENGKW 394

Qy     201 ----WVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAREEEYLFMVGDFLLAVM 256
             ||| :          |:  |||:  : |: |: | |    ||::|   ::: |
Db     395 YLNKWVYNNQG--------NAYIRCFYFTAAVATSTGNNPAPTNVIEYIYMTCSWMMGVF 446

Qy     257 GFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVIDWYQHLQINKKM 316
           || ::| :  :: | |     |     |     |   :  :   || ||: :    :|
Db     447 VFALLLGQIRDIVSNANRNREEFQRKMDLALGECKKLGLKMETTNRVRDWFIYTWQQQKT 506

Qy     317 TNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQPQTYSPGEYVC 376
           :|  ::: || :|: ::|:||| :|||:||:||:|: :|| :|||||:|  : ||: :|
Db     507 LDEKKLIEKLPLKLQTDLALSVHYTTLSKVQLFQDCDRALLRDLVLKLRPVIFLPGDMIC 566

Qy     377 RKGDIGQEMYIIREGQLAVVADDGITQ-YAVLGAGLYFGEISIINIKGNMSGNRRTANIK 435
           |||:|:|||||  :| | ||  |   : :| |  |  |||||:: | ||   |||||:|:
Db     567 LKGDVGKEMYIINQGILQVVGGDHNEKIFAELAQGAVFGEISLLAIGGN---NRRTASIR 623

Qy     436 SLGYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLDVNAEAAEIALQEATES 495
           : ||  || |:|||| :|:  |||||||:  |   :|    | |   |  :|  |   |
Db     624 AKGYCTLFVLAKEDLNDVIRYYPQAQTILRRKAAAMLKNDKKSDEKTE--KIKAQAELED 681

Qy     496 RLRGLDQQLDDLQTKFARLLAELESSALKIAYRI--ERLE----------WQT------- 536
           | :   :|:  | |  | :    |:  ::   ::  |  |          | |
Db     682 RCKINPRQVPKLITLIANMTEMNENKGVQELKKVIEEETEKSRRQSIYYPWSTLQRDDDD 741

Qy     537 -REWPMPEDLAEADDEGEPEEGTSKDEE 563
            ||   |||::  :: : :     |:|
Db     742 EEEWNDEEDLSDVGEDFDLDPTNHSDDE 769
```

```
RESULT  13
Q9W2D5
ID   Q9W2D5      PRELIMINARY;      PRT;  1037 AA.
AC   Q9W2D5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CG17922 PROTEIN.
GN   CG17922.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003455; AAF46757.1; -.
DR   FlyBase; FBgn0034656; CG17922.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR001622; Channel_pore_K.
DR   InterPro; IPR000595; cNMP_binding.
DR   Pfam; PF00027; cNMP_binding; 1.
DR   Pfam; PF00520; ion_trans; 1.
DR   SMART; SM00100; cNMP; 1.
DR   PROSITE; PS00888; CNMP_BINDING_1; 1.
DR   PROSITE; PS00889; CNMP_BINDING_2; 1.
DR   PROSITE; PS50042; CNMP_BINDING_3; 1.
SQ   SEQUENCE   1037 AA;  118393 MW;  FCFB7A1590E7F051 CRC64;
```

```
  Query Match           26.1%;  Score 781.5;  DB 5;  Length 1037;
  Best Local Similarity  31.2%;  Pred. No. 7.6e-50;
```

Matches 193; Conservative 118; Mismatches 226; Indels 81; Gaps 13;

```
Qy     10 TESSPPAPSKARKLLP--------------------------------VLDPSGDYYY 35
Db    407 TKSLHPSPQHKVRLVPAGQSPHRGRLFEADTPRSNVWLCSSLCGANNDERTLDPQGKIYI 466

Qy     36 WWLNTMVFPVMYNLIILVCRACFPDLQHGYLVAWLVLDYTSDLLYLLDMVVRFH-TGFLE 94
Db    467 SWLCVVSLSFLYNAWVIPLRASFPFQTKENTNIWLACDFCADIIYLLDVVFFKHRVMYLF 526

Qy     95 QGILVVDKGRISSRYVRTWSFFLDLASLMPTDVVYVRLGPHTPTLRLNRFLRAPRLFEAF 154
Db    527 EGFWVKNKNLTRKNYMRKLQFKLDLLALLPLELLYFKLGTQAVWLRFPRFFKIQSFWEVF 586

Qy    155 DRTETRTAYPNAFRIAKLMLYIFVVIHWNSCLYFALSRYLGFGRDAWVYPDPAQPGFERL 214
Db    587 RLLDRVISSPHFVRVAKTLTYMLYMIHITAALYYAYSDYQGLGQNRWVFSGKGHP----- 641

Qy    215 RRQYLYSFYFSTLILTTVGDTPPPAREEEYLFMVGDFLLAVMGFATIMGSMSSVI--YNM 272
Db    642 ---YVRCFAFATKTATSIGKNPKPERQGEYVFMTVAWLMGVFVFALLIGQIRDIISTATR 698

Qy    273 NTADAAFYPDHALVKKYMKLQHVNRKLERRVIDWYQHLQINKKMTNEVAILQHLPERLRA 332
Db    699 NKHEYRQLEDETL--EYMRRLNLSQEVQSRVKMWFQFTWEQQRTLDESNILDALPINLKT 756

Qy    333 EVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQPQTYSPGEYVCRKGDIGQEMYIIREGQ 392
Db    757 DIAISVHIQTLSKVQLFADCEEALLRDLVLKLRAVTFLPGDFVCRKGEVGREMYIVKLGQ 816

Qy    393 LAVV---ADDGITQYAVLGAGLYFGEISIINIKGNMSGNRRTANIKSLGYSDLFCLSKED 449
Db    817 VQVMGGPSSDVV--LATLTEGSVFGEISLLGING---ADRRTADVRSKGYSNLFVLSKSD 871

Qy    450 LREVLSEYPQAQTIMEEKGREILLK--MNKLDVNAEAAEIALQ----------EATESRL 497
Db    872 LNEVIAYYPTAQAILKKRARQLMRKNAARERERERARSALQADVVIGNPKTPETAPKL 931

Qy    498 RGLDQQLDDLQTKFARLLAELESSALKIAYRIERLEWQTREWPMPEDLAEADDEGEPEEG 557
Db    932 ---------LQTVIQAL--PFESPAVVLITRGSKRMRRKRQSVQMETIVEPKME---VTG 977

Qy    558 TSKDEEGRASQEGPPGPE 575
Db    978 AAETQDQKGLKTGRCSPD 995
```

RESULT 14
Q93486
ID Q93486 PRELIMINARY; PRT; 695 AA.
AC Q93486; Q93488;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE F14H8.6 PROTEIN.
GN F14H8.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Matthews L.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z81061; CAB02935.4; -.
SQ SEQUENCE 695 AA; 80705 MW; 5C9F516A9C213EBB CRC64;

Query Match 25.0%; Score 748; DB 5; Length 695;
Best Local Similarity 30.3%; Pred. No. 1.4e-47;
Matches 177; Conservative 116; Mismatches 204; Indels 88; Gaps 11;

```
Qy     27 LDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHGYLVAWLVLDYTSDLLYLLDMVV 86
Db    107 VDEHSNTFYLWTG-LVSLVALHPLVFTALSVFQDV-HPFLRNTGPLNLLLDLVNILDLIA 164

Qy     87 RFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLASLMPTDVVYVRLGPHTPTLRLNRFLR 146
Db    165 HTRIEYVENGVAVKNLSKLMHHRLKSTFFILDVIAVVPAESLHI-FGDSFFWSRINRLTK 223

Qy    147 APRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSCLYFALSR---YLGFGRDAWV- 202
Db    224 CYRLFDFSTQTDTRTTSPHAFGLFKLIFICVVIFHWNGCLYFYISKVYNYTTARLEHWIF 283

Qy    203 -YPDPAQP----------------------------------------------GFE 212
Db    284 SYDKIINPILATCVTDIPNHRDFCDTDDLLITHLPENEVQFTVTDYMTTWENKTTTIKFA 343

Qy    213 RLRRQYLYSFYFSTLILTTVGDTPPPAREEEYLFMVGDFLLAVMGFATIMGSMSSVIYNM 272
Db    344 NFFRQYALSFYWSALTLVTLGEQPSPCTTFQNAFEIGDTLLGLVIFAVIVGDVGNMVVAI 403

Qy    273 NTADAAFYPDHAL--VKKYMKLQHVNRKLERRVIDWYQHLQIN-KKMTNEVAILQHLPER 329
Db    404 NLRKSEF--ENVLDGCKRFMVYRKVPNLLRKKAVEYFAYVWAHGGAQVDEEEIAEFLPPR 461

Qy    330 LRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQPQTYSPGEYVCRKGDIGQEMYIIR 389
Db    462 LFGEIAVEIHMDTLKKVKLFESCDPRLLYELILKLQLRVYSPMDYICKKGDVGTEMYIVK 521

Qy    390 EGQLAVVADDGITQYAVLGAGLYFGEISIINIKGNMSGNRRTANIKSLGYSDLFCLSKED 449
Db    522 EGFVEVVSEDGQTIFVTLPAGFVFGELSILNIPGNKNKNLRTASVRSKGYSDLYVLDKED 581

Qy    450 LREVLSEYPQAQTIMEEKGREILLKMNKLDVNAEAAEIALQEATESRLRGLDQQLDDLQT 509
Db    582 LWEALHEYPQAKDSLIQKGIQILEKDKMIDPNM----------------VDDEDESFNG 624

Qy    510 KFARLLAELESSALKIAYRIERLE---------WQTREWPMPEDL 545
Db    625 PIEDYMEHLEHEILKITKMIDQAEDKIHKSNQKMKTRLFGMEMDL 669
```

RESULT 15
O77658
ID O77658 PRELIMINARY; PRT; 938 AA.
AC O77658;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE CYCLIC NUCLEOTIDE-GATED CHANNEL BETA SUBUNIT 1C.
GN CNCBETA.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Wiesner B., Weiner J., Middendorff R., Hagen V., Kaupp U.B.,
RA Weyand I.;
RT "Cyclic nucleotide-gated channels on the flagellum control calcium ion
RT entry into sperm.";
RL J. Cell Biol. 142:0-0(1998).
DR EMBL; AF074012; AAC26127.1; -.
DR InterPro; IPR000636; Cation_chan_non_lig.
DR InterPro; IPR000595; cNMP_binding.
DR Pfam; PF00027; cNMP_binding; 1.
DR Pfam; PF00520; ion_trans; 1.

```
DR   SMART; SM00100; cNMP; 1.
DR   PROSITE; PS00888; CNMP_BINDING_1; 1.
DR   PROSITE; PS00889; CNMP_BINDING_2; 1.
DR   PROSITE; PS50042; CNMP_BINDING_3; 1.
SQ   SEQUENCE   938 AA;  104518 MW;  D1F4FBCF18E53EC4 CRC64;


  Query Match             24.7%;  Score 738;  DB 6;  Length 938;
  Best Local Similarity   30.0%;  Pred. No. 1.2e-46;
  Matches  188;  Conservative  123;  Mismatches  224;  Indels   92;  Gaps   17;

Qy       4 DTKVKTTESSP-PAPSK-----ARKLLPV------------------------------ 26
           |  | : | || |:|:|     | :: |
Db     246 DPDVTSDEESPKPSPAKKAPEPAPEVKPAEAGQVEEEHYCEMLCCKFKRRPWKKYQFPQS 305

Qy      27 LDP-SGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHGYLVAWLVLDYTSDLLYLLDMV 85
           :|| :   |  ||  :|    :|  ::  |  ||      :  ||::||  ||:||||:
Db     306 IDPLTNLMYILWLFFVVLAWNWNCWLIPVRWAFPYQTPDNIHLWLLMDYLCDLIYLLDIT 365

Qy      86 V-RFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLASLMPTDVVYVRLGPHTPTLRLNRF 144
           | :    |:  | :: ||  : : ||::  | :|:  |:| |::|:: | : | ||| |
Db     366 VFQMRLQFVRGGDIITDKKEMRNNYVKSQRFKMDMLCLLPLDLLYLKFGVN-PLLRLPRC 424

Qy     145 LRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSCLYFALSRYLGFGRDAWVYP 204
           |:    ||  :| |:  :    :|: :   |:   :| |||||:  | | | |   |||
Db     425 LKYMAFFEFNNRLESILSKAYVYRVIRTTAYLLYSLHLNSCLYYWASAYEGLGSTHWVY- 483

Qy     205 DPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAREEEYLFMVGDFLLAVMGFATIMGS 264
             : :   |:  :|::   | |:|  | |    | :|   ::   |  |: ::|
Db     484 -------DGVGNSYIRCYYWAVKTLITIGGLPDPRTLFEIVFQGLNYFTGVFAFSVMIGQ 536

Qy     265 MSSVIYNMNTADAAFY---PDHALVKKYMKLQHVNRKLERRVIDWYQHLQINKKMTNEVA 321
           |  |:    ||   :|    |  :  |||    : | :: ||  ||::   :: | :|
Db     537 MRDVV-GAATAGQTYYRSCMDSTV--KYMNFYKIPRSVQNRVKTWYEYTWHSQGMLDESE 593

Qy     322 ILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQPQTYSPGEYVCRKGDI 381
           ::  ||:::| ::|: |: | :|:| :|| |:  :: ::: :|:   | | :|||:||:|
Db     594 LMVQLPDKMRLDLAIDVNYSIVSKVALFQGCDRQMIFDMLKRLRSVVYLPNDYVCKKGEI 653

Qy     382 GQEMYIIREGQLAVV-ADDGITQYAVLGAGLYFGEISIINIKGNMSGNRRTANIKSLGYS 440
           |:|||||: ||: |:   || :    | ||  |||||:: : |   |||||||: : |::
Db     654 GREMYIIQAGQVQVLGGPDGKSVLVTLKAGSVFGEISLLAVGG---GNRRTANVVAHGFT 710

Qy     441 DLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNK--------LDVNAEAAEI---AL 489
           :|| | |:|| |:|  ||::| :: :| | :|   ||        |   |   ::   ||
Db     711 NLFILDKKDLNEILVHYPESQKLLRKKARRMLRNNNKPKEKSVLILPPRAGTPKLFNAAL 770

Qy     490 QEATESRLRGLDQQLDDLQTKFARLLAEL-ESSALKIAYRIERLEWQTREWPMPEDLAEA 548
             | :   :|          : | | | | | :||: | | ::|
Db     771 AAAGKMGAKG------GRGGRLALLRARLKELAALEAAARQQQLL--------------- 809

Qy     549 DDEGEPEEGTSKDEEGRASQEGPPGPE 575
            :: :  |  :  ||| || | || ||
Db     810 -EQAKSSEDAAVGEEGSASPEQPPRPE 835


Search completed: June 20, 2002, 17:15:03
Job time: 531 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 20, 2002, 14:28:47 ; Search time 61.45 Seconds
(without alignments)
1039.340 Million cell updates/sec

Title: US-09-927-267-1
Perfect score: 2989
Sequence: 1 MSQDTKVKTTESSPPAPSKA.....EGTSKDEEGRASQEGPPGPE 575

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1535	51.4	664	22	AAE04894	Human transporter
2	1235	41.3	665	22	ABB63493	Drosophila melanog
3	947	31.7	702	22	ABB65859	Drosophila melanog
4	829.5	27.8	1693	22	ABB64282	Drosophila melanog
5	808	27.0	201	22	AAM24179	Rat EST encoded pr
6	781.5	26.1	1037	22	ABB66545	Drosophila melanog
7	714	23.9	652	22	ABG27471	Novel human diagno
8	571	19.1	259	22	ABG27071	Novel human diagno
9	540	18.1	381	22	ABG15758	Novel human diagno
10	540	18.1	806	22	ABG13481	Novel human diagno
11	488	16.3	418	22	ABB59972	Drosophila melanog
12	467.5	15.6	1203	22	AAB86462	Human HCN4 protein
13	467.5	15.6	1245	22	ABB11953	Human cation chann
14	466.5	15.6	528	20	AAY22193	Mouse brain CNG-2
15	465.5	15.6	889	22	AAB86461	Human HCN2 protein
16	464.5	15.5	863	22	AAB86463	Murine HCN2 protei
17	456.5	15.3	774	21	AAB28375	Human hyperpolaris
18	448.5	15.0	598	20	AAY22194	Human brain CNG-2
19	445	14.9	767	20	AAY31227	Sea urchin Ih ion
20	441.5	14.8	755	20	AAY22192	Human brain CNG-1
21	441.5	14.8	910	20	AAY22191	Mouse brain CNG-1
22	422	14.1	1307	22	ABB63936	Drosophila melanog
23	406.5	13.6	506	20	AAY22195	Mouse brain CNG-3
24	391	13.1	962	21	AAY49944	Human potassium io
25	386.5	12.9	1174	22	ABB65304	Drosophila melanog
26	382.5	12.8	962	21	AAY83032	Rat Eag1 potassium
27	382.5	12.8	1311	22	ABB61234	Drosophila melanog
28	380.5	12.7	989	21	AAY49945	Human potassium io
29	380	12.7	1080	21	AAY77738	Human ESK1 (hESK1)
30	380	12.7	1083	20	AAY22426	Human brain specif
31	380	12.7	1083	21	AAY44778	Human Elk voltage
32	380	12.7	1083	21	AAY44779	Human Elk voltage
33	380	12.7	1083	21	AAY44780	Human Elk voltage
34	380	12.7	1083	21	AAY44781	Human Elk voltage
35	379	12.7	1083	21	AAY44904	Monkey potassium c
36	377	12.6	1082	20	AAY34128	Human potassium ch
37	377	12.6	1083	21	AAY84835	Herg4, a potassium
38	376.5	12.6	988	22	AAE01043	Human ether a gogo
39	376.5	12.6	988	22	AAB31714	A human alpha-subu
40	376.5	12.6	1107	21	AAY44907	Human potassium ch
41	375	12.5	1159	20	AAY32020	Human cation chann
42	375	12.5	1159	21	AAY85405	Long QT syndrome a
43	375	12.5	1159	21	AAY85406	Long QT syndrome a
44	375	12.5	1159	22	AAM51812	Human potassium ch
45	375	12.5	1159	22	AAB70939	Human eag-related

ALIGNMENTS

RESULT 1
AAE04894
ID AAE04894 standard; Protein; 664 AA.
XX
AC AAE04894;
XX
DT 10-SEP-2001 (first entry)
XX
DE Human transporter and ion channel-7 (TRICH-7) protein.
XX
KW Human; transporter and ion channel-7; TRICH-7; vaccine; cystic fibrosis;
KW gene therapy; amyotrophic lateral sclerosis; amnesia; muscular dystrophy;
KW hypertension; angina; neurological disorder; asthma; bipolar disorder;
KW dementia; depression; Alzheimer's disease; epilepsy; mood; arrhythmia;
KW Pick's disease; ischaemic cerebrovascular disease; AIDS; anxiety; stroke;
KW Huntington's disease; Parkinson's disease; cerebral neoplasm; allergy;
KW demyelinating disease; mental disorder; Schizophrenia; polymyositis;
KW muscle disorder; cardiomyopathy; cataract; myocarditis; Grave's disease;
KW dermatomyositis; diabetes mellitus; immunological disorder; psoriasis;
KW rheumatoid arthritis; Sjogren's syndrome; systemic lupus erythematosus;
KW sickle cell anaemia; Wilson's disease; infertility; Cushing's disease;
KW scleroderma; pulmonary artery stenosis; nootropic; Addison's disease;
KW malabsorption syndrome; hypercholesterolaemia; cancer.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 460..476
FT /label= Transmembrane_domain
FT Domain 469..565
FT /note= "Cyclic nucleotide binding domain"
FT Domain 478..501
FT /label= Transmembrane_domain
FT Domain 516..525

```
FT                   /label= Transmembrane_domain
XX
PN   WO200146258-A2.
XX
PD   28-JUN-2001.
XX
PF   22-DEC-2000; 2000WO-US35095.
XX
PR   23-DEC-1999;   99US-0172000.
PR   14-JAN-2000; 2000US-0176083.
PR   21-JAN-2000; 2000US-0177332.
PR   28-JAN-2000; 2000US-0178572.
PR   02-FEB-2000; 2000US-0179758.
PR   10-FEB-2000; 2000US-0181625.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Baughn MR,  Burford N,  Au-Young J,  Lu DAM,  Yang J,  Reddy R,  Lal P;
PI   Hillman JL,  Azimzai Y,  Yue H,  Nguyen DB,  Yao MG,  Gandhi AR;
PI   Tang YT,  Khan FA;
XX
DR   WPI; 2001-418042/44.
DR   N-PSDB; AAD09558.
XX
PT   Novel human transporter and ion channel proteins useful for treating
PT   and preventing transport, neurological, muscle and immunological
PT   disorders -
XX
PS   Claim 1; Page 119-120; 160pp; English.
XX
CC   The present sequence is transporter and ion channel-7 (TRICH-7) protein.
CC   TRICH is used as vaccine. TRICH is useful for treating a disease or
CC   condition associated with decreased expression of functional TRICH,
CC   such as transport disorder including amyotrophic lateral sclerosis,
CC   cystic fibrosis, Becker's muscular dystrophy, Charcot-Marie Tooth
CC   disease, Duchenne muscular dystrophy, angina and hypertension,
CC   neurological disorders including Alzheimer's disease, amnesia, bipolar
CC   disorder, dementia, depression, epilepsy, ischaemic cerebrovascular
CC   disease, stroke, cerebral neoplasms, Pick's disease, Huntington's
CC   disease and Parkinson's disease, demyelinating diseases, mental disorders
CC   including mood, anxiety, Schizophrenia and seasonal affective disorder,
CC   muscle disorder including cardiomyopathy, myocarditis, polymyositis,
CC   dermatomyositis, arrhythmias and asthma and immunological disorders
CC   including AIDS, adult respiratory distress syndrome (ARDS), allergies,
CC   anaemia, diabetes mellitus, rheumatoid arthritis, scleroderma, Sjogren's
CC   syndrome, systemic lupus erythematosus and other diseases including
CC   sickle cell anaemia, Wilson's disease, cataracts, infertility, pulmonary
CC   artery stenosis, Grave's disease, Cushing's disease, Addison's disease,
CC   glucose-galactose malabsorption syndrome, hypercholesterolaemia, cancers
CC   psoriasis and viral, bacterial, fungal, helminthic and protozoal
CC   infections. TRICH DNA is useful in gene therapy and in diagnostic
CC   purposes.
XX
SQ   Sequence   664 AA;


  Query Match           51.4%;  Score 1535;  DB 22;  Length 664;
  Best Local Similarity   53.9%;  Pred. No. 2.3e-138;
  Matches  291;  Conservative  104;  Mismatches  137;  Indels    8;  Gaps    3;

Qy      18 SKARKLLPVLDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHGYLVAWLVLDYTSD 77
           :| :  | ||||:||:|| ||  :  ||:||  :|| |||| ||| || : ||||||  ||
Db     124 tkkkfelfvldpagdwyycwlfviampvlynwcllvaracfsdlqkgyylvwlvldyvsd 183

Qy      78 LLYLLDMVVRFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLASLMPTDVVYVRLGPHTP 137
           ::|: |: :|  |||||||:|| |  ::   |: |  | ||:||::|||::|  :  |:|
Db     184 vvyiadlfirlrtgfleqgllvkdtkklrdnyihtlqfkldvasiiptdliyfavdihsp 243

Qy     138 TLRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSCLYFALSRYLGFG 197
           :| || |   |:|| |||||||| ||| |||: |:||| |:||||:|:|:|:|: :|||
Db     244 evrfnrllhfarmfeffdrtetrtnypnifrisnlvlyilviihwnaciyyaisksigfg 303

Qy     198 RDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAREEEYLFMVGDFLLAVMG 257
           | ||||:   | :  | |:|:|  |:||| |||:|:|||| ::|||||:: |||: |:
Db     304 vdtwvypnitdpeygylareyiyclywstltlttigetpppvkdeeylfvifdfligvli 363

Qy     258 FATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVIDWYQHLQINKKMT 317
           ||||:|:: |:| ||| | |      || ||: : |:: :| :|| |: :|  |||
Db     364 fativgnvgsmisnmnatraefqakidavkhymqfrkvskgmeakvirwfdylwtnkktv 423

Qy     318 NEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQPQTYSPGEYVCR 377
           :|  ||::|| :||||:|::|||||| :|:|| :||| || ||||||:|| :|||:|:||
Db     424 dereilknlpaklraeiainvhlstlkkvrifhdceagllvelvlklrpqvfspgdyicr 483

Qy     378 KGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIINIKGNMSGNRRTANIKSL 437
           |||||:|||||:||:|||||||||:||||:| ||  ||||||:||||:  ||||||||:||
Db     484 kgdigkemyiikegklavvaddgvtqyallsagscfgeisilnikgskmgnrrtanirsl 543

Qy     438 GYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLDVN--AEAAEIALQEATES 495
           ||||||||||:|| | ::||| |: ::||:|||||:|   || |  | : |: :||
Db     544 gysdlfclskddlmeavteypdakkvleergreilmkeglldenevatsmevdvqe---- 599

Qy     496 RLRGLDQQLDDLQTKFARLLAELESSALKIAYRIERLEWQTREWPMPEDLAEADDEGEPE 555
           :|  |:  :: | |:| |||||   :  |:  ||  || : ::    ||   :|    ||
Db     600 klgqletnmetlytrfgrllaeytgaqqklkqritvletkmkq--nneddylsdgmnspe 657


RESULT   2
ABB63493
ID   ABB63493 standard; Protein; 665 AA.
XX
AC   ABB63493;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 17271.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL07596.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 17271; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
```

```
SQ   Sequence   665 AA;

  Query Match             41.3%;  Score 1235;  DB 22;  Length 665;
  Best Local Similarity   47.1%;  Pred. No. 1.6e-109;
  Matches  243;  Conservative   98;  Mismatches  165;  Indels   10;  Gaps    2;

Qy      27 LDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHGYLVAWLVLDYTSDLLYLLDMVV 86
           :||:   :| ||  :   |:||:| :| || | ::       |  |||  | :|||| :|
Db     100 vdptlqshyrwlaivslavlyniifvvgravfweinksapafwytldylcdfiylldtlv 159

Qy      87 RFHTGFLEQGILVVDKGRISSRYVRTWSFFLDLASLMPTDVVYVRLGPHTPT-------- 138
             | |||:||:|| |  |:   |  |  ::||: |::|||: |:   | | :
Db     160 hmhegfldqgllvrdafrlrrhyfhtkgwyldvlsmlptdlayiwwppetcsslylpcpv 219

Qy     139 -LRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSCLYFALSRYLGFG 197
            :|||| ||  ||:| |||||| | ||||||| |::| | |:||||:|:|||:|  :||
Db     220 ivrlnrllrinrlwewfdrtetatgypnafrickvvlailvlihwnacmyfaisyeigfs 279

Qy     198 RDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAREEEYLFMVGDFLLAVMG 257
            |:|||          |:|||:||||:||| |||:|:|| |  : ||||:| |||  |:
Db     280 sdswvynlngtrn-ntlqrqyiysfywstltlttigetptpendveylfvvadflagvli 338

Qy     258 FATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVIDWYQHLQINKKMT 317
           ||||:|:: |:| ||| |   |      ||:||  : |  :|| ||| |: :
Db     339 fativgnigsmisnmnvarvefqnrmdgvkqymafrrvgheleavrivrfaytwsqsgal 398

Qy     318 NEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQPQTYSPGEYVCR 377
           :|  :|  ||::|:||:|: ||: || :|:|| : |  ||| |||||: | :|||:|:||
Db     399 deervlaalpdklkaeiaiqvhmdtlkqvrifhdtepgllealvlklklqvfspgdyicr 458

Qy     378 KGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIINIKGNMSGNRRTANIKSL 437
           |||:|:|||||:: |:|:|| |||||  | ||||  |||:|:: | || :|||||||::||
Db     459 kgdvgkemyivkrgklsvvgddgitvlatlgagsvfgevsvleiagnrtgnrrtanvrsl 518

Qy     438 GYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLDVNAEAAEIALQEATESRL 497
           |||||||:| || | ||:||:|:: : ::| ::| |   ||    |    : :: | :
Db     519 gysdlfclakrdlwetlsdypearstltqrgcqllrkdglldeqifadsqrvhdsieggi 578

Qy     498 RGLDQQLDDLQTKFARLLAELESSALKIAYRIERLE 533
             |:  :::|  : ||||||  :|  ||  |: :||
Db     579 eklelsvenlnmrlarllaeytasqakikqrlakle 614


RESULT   3
ABB65859
ID   ABB65859 standard; Protein; 702 AA.
XX
AC   ABB65859;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 24369.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL09962.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 24369; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   702 AA;

  Query Match             31.7%;  Score 947;  DB 22;  Length 702;
  Best Local Similarity   42.2%;  Pred. No. 8.9e-82;
  Matches  185;  Conservative   95;  Mismatches  150;  Indels    8;  Gaps    4;

Qy      40 TMVFPVMYNLIILVCRACFPDLQHGYLVAWLVLDYTSDLLYLLDMVVRFHTGFLEQGILV 99
           :| |  :||  ::: |  | ::     :  |  ||| || |||:|::  | ||:|| |:|
Db       4 smaf--lynfwviiyrfafqeinrrtiaiwfcldylsdflylidilfhfrtgyledgvlq 61

Qy     100 VDKGRISSRYVRTWSFFLDLASLMPTDVVYVRLGPHTPTLRLNRFLRAPRLFEAFDRTET 159
             |  :: : |: :  |::|   |:| | :|: :|     ||  | ::  | :   ||||
Db      62 tdalklrthymnstifyidclcllpldflylsig-fnsilrsfrlvkiyrfwafmdrter 120

Qy     160 RTAYPNAFRIAKLMLYIFVVIHWNSCLYFALSRYLGFGRDAWVYPDPAQPGFERLRRQYL 219
             | ||| ||   |: |: |: ||| |||  : :  |||   ||| |          : :|||
Db     121 htnypnlfrstalihyllvifhwngclyhiihknngfgsrnwvyhdsesad---vvkqyl 177

Qy     220 YSFYFSTLILTTVGDTPPPAREEEYLFMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAF 279
           |:|: || |||:|| | |  : ||:|::   |  :| |||::| ::::: ::: |   |
Db     178 qsyywctlalttigdlpkprskgeyvfvilqllfglmlfatvlghvanivtsvsaarkef 237

Qy     280 YPDHALVKKYMKLQHVNRKLERRVIDWYQHLQINKKMTNEVAILQHLPERLRAEVAVSVH 339
                || ||::: |   |: :|| |: :| : :| ::|   :  ||::|:||:|::||
Db     238 qakldgvktymrmrrvpnhlqvkvikwfdylwltqkcsdeeravsclpdklkaeiainvh 297

Qy     340 LSTLSRVQIFQNCEASLLEELVLKLQPQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADD 399
           | || ||:|||| ||  | ||||:|:|  :|||:|:||||::|:||||:  |:| ||||:
Db     298 ldtlkrveifqnteagflcelvlrlrpvlfspgdyicrkgevgkemyivnrgrlqvvadn 357

Qy     400 GITQYAVLGAGLYFGEISIINIKGNMSGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQ 459
           | |  | | || |||||||:|:    :||||||:::|:|||||| |||:|: :|| |||
Db     358 gktvmaslkagsyfgeisilnm--gtagnrrtasvrsvgysdlfvlskkdmwdvlkeypa 415

Qy     460 AQTIMEEKGREILLKMNK 477
           |:  :|    : | |  |
Db     416 arvrlesiavkrlekykk 433


RESULT   4
ABB64282
ID   ABB64282 standard; Protein; 1693 AA.
XX
AC   ABB64282;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 19638.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
```

```
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL08385.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 19638; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1693 AA;
```

```
  Query Match             27.8%;  Score 829.5;  DB 22;  Length 1693;
  Best Local Similarity   33.3%;  Pred. No. 7e-70;
  Matches  197;  Conservative  116;  Mismatches  204;  Indels   75;  Gaps   14;

Qy      45 VMYNLIILVCRACFPDLQHGYLVAWLVLDYTSDLLYLLDMVVRFHTGFLEQGILVVDKGR 104
           |:|||  |: |  ||:||      ||: |  :|::::||::|:  ||:||||::| |  :
Db       7 vlynlwtlivrqsfpelqqsvptfwlicdsmtdvvfildiivqlrtgyleqglmvyddrk 66

Qy     105 ISSRYVRTWSFFLDLASLMPTDVVYVRLGPHTPTLRLNRFLRAPRLFEAFDRTETRTAYP 164
           ::  || :  |  |: :|:| |:: :::| | | ||  |  |: | :      |:|| :|
Db      67 lachyvhsrdfifdmialipldllqlkmgth-pllrftrvyrsvrfyyi---vesrtvwp 122

Qy     165 NAFRIAKLMLYIFVVIHWNSCLYFALSRYLGFGRDAWVYPDPAQPG-FERLRRQYLYSFY 223
           | :|:  |:  : :: ||  | || ||   ||  | |||  | :|| :  | |:|| | |
Db     123 nlwrvvnlihillilahwfgcfyfllseaegfqgd-wvy--pyrpgdyatltrkylgsly 179

Qy     224 FSTLILTTVGDTPPPAREEEYLFMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDH 283
           :||| |||:|| | |    ||:| :  :|: |  ||||:| : :|| | |     |
Db     180 wstltlttigdlptpetnaeyiftivsyligvfifativgqvgnvitnrnanrleferll 239

Qy     284 ALVKKYMKLQHVNRKLERRVIDWYQHLQINKKMTNEVAI---LQHLPERLRAEVAVSVHL 340
           | ||:   |   ::|||: || :     ::     |   |  ||::|: |:|: |:|
Db     240 dgaktymrhhkvpggmkrrvlrwydyswsrgriqgggdintalgllpdklktelalhvnl 299

Qy     341 STLSRVQIFQNCEASLLEELVLKLQPQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADDG 400
           | | :| ||| |:   | :||||::   ::||: :||||:: :||:|| :| | |::: |
Db     300 svlkkvtifqecqpeflhdlvlkmkayiftpgdsicrkgevaremfiiadgilevlsetg 359

Qy     401 ITQYAVLGAGLYFGEISIINIKGNMSGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQA 460
                : || :|||| |:|: |    |:|||:::|:|||:|| ||:||:   : :|| |
Db     360 -kvlttmkagdffgeigilnldgl---nkrtadvrsvgyselfslsredvlaamkdypda 415

Qy     461 QTIMEEKGREILLK---MNKLDVNAEAAEIALQEAT------ESRLRGLDQQLDDLQTKF 511
           | |::  ||: |::   :||    |:: : |   |       :|  :|   | |  :
Db     416 qeilqtlgrkrlmevrcvnkkyakaqsdkeaaayaaahphhhqshhqgqvhqsdssensa 475

Qy     512 ARLLAELESSALKIAYRI----------ERLEWQTREWPMPE---------------- 543
           :: : :    :|   :           | || |      |
Db     476 skkivdklkhdvkgfrnvlkksrtsrksdeslemqplhntsprgskillkrmsrvrsdek 535

Qy     544 --DLAEADDE----------GEP----------EEGTSKDEEGRASQEGPP 572
             | ||| ||          | |          :: : ::|  |   || |
Db     536 dadsaeakdelhdktpspigaglpllqrlrllkekqesiqeeperefsegfp 587
```

```
RESULT   5
AAM24179
ID   AAM24179 standard; Protein; 201 AA.
XX
AC   AAM24179;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Rat EST encoded protein SEQ ID NO: 1704.
XX
KW   Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW   tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW   diagnostics; forensic test; gene mapping; genetic disorder;
KW   biodiversity; gene therapy; nutrition.
XX
OS   Rattus norvegicus.
XX
PN   WO200154477-A2.
XX
PD   02-AUG-2001.
XX
PF   25-JAN-2001; 2001WO-US02687.
XX
PR   25-JAN-2000; 2000US-0491404.
PR   17-JUL-2000; 2000US-0617746.
PR   03-AUG-2000; 2000US-0631451.
PR   15-SEP-2000; 2000US-0663870.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Zhou P,  Qian XB,  Wang Z,  Chen R,  Asundi V;
PI   Cao Y,  Drmanac RA,  Zhang J,  Werhman T;
XX
DR   WPI; 2001-476164/51.
DR   N-PSDB; AAH98838.
XX
PT   Isolated polypeptide for treatment of diseases, diagnostics, raising
PT   antibodies and research use -
XX
PS   Claim 20; Page 1136; 1275pp; English.
XX
CC   The present invention provides the protein and coding sequences of novel
CC   proteins from a variety of organisms, including human, dog, cat, horse,
CC   cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC   urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC   from the organism of interest. They can be used in diagnostics,
CC   forensics, gene mapping, identification of mutations, to assess
CC   biodiversity and for nutritional purposes. The present sequence is a
CC   protein of the invention.
XX
SQ   Sequence   201 AA;
```

```
  Query Match             27.0%;  Score 808;  DB 22;  Length 201;
  Best Local Similarity   91.1%;  Pred. No. 3e-69;
  Matches  154;  Conservative    3;  Mismatches   12;  Indels    0;  Gaps    0;

Qy      41 MVFPVMYNLIILVCRACFPDLQHGYLVAWLVLDYTSDLLYLLDMVVRFHTGFLEQGILVV 100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 mvfpvmynliilvcracfpdlqhgylvawlvldytsdllylldmvvrfhtgfleqgilvv 60

Qy     101 DKGRISSRYVRTWSFFLDLASLMPTDVVYVRLGPHTPTLRLNRFLRAPRLFEAFDRTETR 160
```

```
Db     61 dkgrissryvrtwsffldlaslmptdvvyvrlgphtptlrlnrflraprlfeafdrtetr 120

Qy    161 TAYPNAFRIAKLMLYIFVVIHWNSCLYFALSRYLGFGRDAWVYPDPAQP 209
Db    121 taypnafcigklmlyifgrihwnnclyfslsrylgfgrepmgvprtpap 169


RESULT   6
ABB66545
ID   ABB66545 standard; Protein; 1037 AA.
XX
AC   ABB66545;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 26427.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL10648.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 26427; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1037 AA;


  Query Match           26.1%;  Score 781.5;  DB 22;  Length 1037;
  Best Local Similarity   31.2%;  Pred. No. 1.3e-65;
  Matches  193;  Conservative  118;  Mismatches  226;  Indels   81;  Gaps   13;

Qy     10 TESSPPAPSKARKLLP----------------------------------VLDPSGDYYY 35
Db    407 tkslhpspqhkvrlvpagqsphrgrlfeadtprsnvwlcsslcganndertldpqgkiyi 466

Qy     36 WWLNTMVFPVMYNLIILVCRACFPDLQHGYLVAWLVLDYTSDLLYLLDMVVRFH-TGFLE 94
Db    467 swlcvvslsflynawviplrasfpfqtkentniwlacdfcadiiylldvvffkhrvmylf 526

Qy     95 QGILVVDKGRISSRYVRTWSFFLDLASLMPTDVVYVRLGPHTPTLRLNRFLRAPRLFEAF 154
Db    527 egfwvknknltrknymrklqfkldllallplellyfklgtqavwlrfprffkiqsfwevf 586

Qy    155 DRTETRTAYPNAFRIAKLMLYIFVVIHWNSCLYFALSRYLGFGRDAWVYPDPAQPGFERL 214
Db    587 rlldrvissphfvrvaktltymlymihitaalyyaysdyqglgqnrwvfsgkghp----- 641

Qy    215 RRQYLYSFYFSTLILTTVGDTPPPAREEEYLFMVGDFLLAVMGFATIMGSMSSVI--YNM 272
Db    642 ---yvrcfafatktatsigknpkperqgeyvfmtvawlmgvfvfalligqirdiistatr 698

Qy    273 NTADAAFYPDHALVKKYMKLQHVNRKLERRVIDWYQHLQINKKMTNEVAILQHLPERLRA 332
Db    699 nkheyrqledetl--eymrrlnlsqevqsrvkmwfqftweqqrtldesnildalpinlkt 756

Qy    333 EVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQPQTYSPGEYVCRKGDIGQEMYIIREGQ 392
Db    757 diaisvhiqtlskvqlfadceeallrdlvlklravtflpgdfvcrkgevgremyivklgq 816

Qy    393 LAVV---ADDGITQYAVLGAGLYFGEISIINIKGNMSGNRRTANIKSLGYSDLFCLSKED 449
Db    817 vqvmggpssdvv--latltegsvfgeisllging---adrrtadvrskgysnlfvlsksd 871

Qy    450 LREVLSEYPQAQTIMEEKGREILLK--MNKLDVNAEAAEIALQ----------EATESRL 497
Db    872 lneviayyptaqailkkrarqlmrknaarereeererarsalqadvvignpktpetapkl 931

Qy    498 RGLDQQLDDLQTKFARLLAELESSALKIAYRIERLEWQTREWPMPEDLAEADDEGEPEEG 557
Db    932 ---------lqtviqal--pfespavvlitrgskrmrrkrqsvqmetivepkme---vtg 977

Qy    558 TSKDEEGRASQEGPPGPE 575
Db    978 aaetqdqkglktgrcspd 995


RESULT   7
ABG27471
ID   ABG27471 standard; Protein; 652 AA.
XX
AC   ABG27471;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #27462.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS91658.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 20; SEQ ID No 57830; 103pp; English.
```

```
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   652 AA;


  Query Match             23.9%;  Score 714;  DB 22;  Length 652;
  Best Local Similarity   30.9%;  Pred. No. 2e-59;
  Matches  174;  Conservative  120;  Mismatches  215;  Indels   54;  Gaps   15;

Qy      34 YYWWLNTMVFPVMYNLIILVCRACFPDLQHGYLVAWLVLDYTSDLLYLLDMVVRFHT--G 91
           |  ||  :|    :|  ::  |  ||       :  ||::||  ||:| ||: | | |
Db      57 yvlwlffvvmawnwncwlipvrwafpyqtpdnihhwllmdylcdliyflditv-fqtrlq 115

Qy      92 FLEQGILVVDKGRISSRYVRTWSFFLDLASLMPTDVVYVRLGPHTPTLRLNRFLRAPRLF 151
           |:  | :: ||  : : |:::  | :|| ||:| | :|:::| : | ||| | |:    |
Db     116 fvrggdiitdkkdmrnnylksrrfkmdllsllpldflylkvgvn-pllrlprclkymaff 174

Qy     152 EAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSCLYFALSRYLGFGRDAWVYPDPAQPGF 211
           |   | |:  :    :|: :   |:   :| |||||:  | | | |   |||
Db     175 efnsrlesilskayvyrvirttayllyslhlnsclyywasayqglgsthwvy-------- 226

Qy     212 ERLRRQYLYSFYFSTLILTTVGDTPPPAREEEYLFMVGDFLLAVMGFATIMGSMSSVIYN 271
           : :   |:  :||:   | |:|  | |    | :| : ::   |  |: ::| |  |:
Db     227 dgvgnsyircyyfavktlitigglpdpktlfeivfqllnyftgvfafsvmigqmrdvv-g 285

Qy     272 MNTADAAFY---PDHALVKKYMKLQHVNRKLERRVIDWYQHLQINKKMTNEVAILQHLPE 328
             ||   :|    |  :  |||    : : :: ||  ||::   :: | :|  ::  ||:
Db     286 aatagqtyyrscmdstv--kymnfykipksvqnrvktwyeytwhsqgmldeselmvqlpd 343

Qy     329 RLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQPQTYSPGEYVCRKGDIGQEMYII 388
           ::| ::|: |: : :|:| :|| |:  :: ::: :|:   | | :|||:||:||:|||||
Db     344 kmrldlaidvnynivskvalfqgcdrqmifdmlkrlrsvvylpndyvckkgeigremyii 403

Qy     389 REGQLAVV-ADDGITQYAVLGAGLYFGEISIINIKGNMSGNRRTANIKSLGYSDLFCLSK 447
           : ||: |:   || :    | ||  |||||:: : |   |||||||: : |:::|| | |
Db     404 qagqvqvlggpdgksvlvtlkagsvfgeisllavgg---gnrrtanvvahgftnlfildk 460

Qy     448 EDLREVLSEYPQAQTIMEEKGREILLKMNKLDVNAEAAEIALQE----------ATES 495
           :|| |:|  ||::| :: :| | :|   ||     | : : |           |
Db     461 kdlneilvhypesqkllrkkarrmlrsnnk--pkeeksvlilppragtpklfnaalamtg 518

Qy     496 RLRGLDQQLDDLQTKFARLLAEL-ESSALKIAYRIERLEWQTREWPMPEDLAEADDEGEP 554
           :: |   :      | | | | | :||: | : | |  |          |::  : :
Db     519 kmggkgak----ggklahlrarlkelaaleaaakheelveq---------akssqdvkg 564

Qy     555 EEGTSKDEEGRASQEG---PPGP 574
           |||::  ::    :|    || |
Db     565 eegsaapdqhthpkeaatdppap 587


RESULT   8
ABG27071
```

```
ID   ABG27071 standard; Protein; 259 AA.
XX
AC   ABG27071;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #27062.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS91258.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 57430; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   259 AA;


  Query Match             19.1%;  Score 571;  DB 22;  Length 259;
  Best Local Similarity   50.7%;  Pred. No. 2.7e-46;
  Matches  103;  Conservative   46;  Mismatches   54;  Indels    0;  Gaps    0;

Qy      92 FLEQGILVVDKGRISSRYVRTWSFFLDLASLMPTDVVYVRLGPHTPTLRLNRFLRAPRLF 151
           :| ||:||  : :: ::|     | ||: ||:|||::| :|| : | :|||| ||  |:|
Db      57 ylnqgllvkkelklinkyksnlqfkldvlsliptdllyfklgwnypeirlnrllrfsrmf 116

Qy     152 EAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSCLYFALSRYLGFGRDAWVYPDPAQPGF 211
           | | |||||| ||| |||: |::|| ::|||':|:::::|: :||| | |||||   | |
Db     117 effqrtetrtnypnifrisnlvmyiviiihwnacvfysiskaigfgndtwvypdindpef 176
```

```
Qy    212 ERLRRQYLYSFYFSTLILTTVGDTPPPAREEEYLFMVGDFLLAVMGFATIMGSMSSVIYN 271
          || |:|:|| |:||| |||:|:|||| |: ||:|:| |||: |: ||||:|:: |:| |
Db    177 grlarkyvyslywstltlttigetpppvrdseyvfvvvdfligvlifativgnigsmisn 236

Qy    272 MNTADAAFYPDHALVKKYMKLQH 294
          || | | |      :|:||  ::
Db    237 mnaaraefqaridaikqymhfrn 259


RESULT   9
ABG15758
ID   ABG15758 standard; Protein; 381 AA.
XX
AC   ABG15758;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #15749.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS79945.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 46117; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   381 AA;


  Query Match             18.1%;  Score 540;  DB 22;  Length 381;
  Best Local Similarity   48.1%;  Pred. No. 4.8e-43;
  Matches  104;  Conservative   44;  Mismatches   32;  Indels   36;  Gaps    2;

Qy    241 EEEYLFMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMK--------- 291
          : ||:|:| |||: |: |||||:|:: |:| ||| | | |      :|:|::
Db      1 dseyvfvvvdfligvlifativgnigsmisnmnaaraefqaridaikqyvrgrnqdlmls 60

Qy    292 -----LQHVNRKLERRVIDWYQHLQINKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRV 346
               ::|:: :|:||| |: :|  ||| :|  :|::||::||||:|::||| |: :|
Db     61 snictfrnvskdmekrvikwfdylwtnkktvdekevlkylpdklraeiainvhldtvgkv 120

Qy    347 QIFQNCEASLLEELVLKLQPQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADDGITQYAV 406
                                 ||||:|:|:|||||:|||||:||:|||||||||:||: |
Db    121 ----------------------yspgdyickkgdigremyiikegklavvaddgvtqfvv 158

Qy    407 LGAGLYFGEISIINIKGNMSGNRRTANIKSLGYSDL 442
          |  | |||||||:||||: :||||||||||:|||||
Db    159 lsdgsyfgeisilnikgskagnrrtaniksigysdl 194


RESULT  10
ABG13481
ID   ABG13481 standard; Protein; 806 AA.
XX
AC   ABG13481;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #13472.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS77668.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 43840; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
```

```
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   806 AA;


  Query Match             18.1%;  Score 540;  DB 22;  Length 806;
  Best Local Similarity   48.1%;  Pred. No. 1.5e-42;
  Matches  104;  Conservative   44;  Mismatches   32;  Indels   36;  Gaps    2;

Qy     241 EEEYLFMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMK--------- 291
           : ||:|:| |||: |: ||||:|:: |:| ||| | | |      :|:|::
Db       1 dseyvfvvvdfligvlifativgnigsmisnmnaaraefqaridaikqyvrgrnqdlmls 60

Qy     292 -----LQHVNRKLERRVIDWYQHLQINKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRV 346
                ::|:: :|:||| |: :|  |||  :|  :|::||::||||:|::||| |: :|
Db      61 snictfrnvskdmekrvikwfdylwtnkktvdekevlkylpdklraeiainvhldtvgkv 120

Qy     347 QIFQNCEASLLEELVLKLQPQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADDGITQYAV 406
                              ||||:|:|:|||||:|||||:||:|||||||||:||: |
Db     121 ----------------------yspgdyickkgdigremyiikegklavvaddgvtqfvv 158

Qy     407 LGAGLYFGEISIINIKGNMSGNRRTANIKSLGYSDL 442
           |  | |||||||:||||: :||||||||||:|||||
Db     159 lsdgsyfgeisilnikgskagnrrtaniksigysdl 194


RESULT  11
ABB59972
ID   ABB59972 standard; Protein; 418 AA.
XX
AC   ABB59972;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 6708.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL04075.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 6708; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   418 AA;


  Query Match             16.3%;  Score 488;  DB 22;  Length 418;
  Best Local Similarity   50.5%;  Pred. No. 5.5e-38;
  Matches   95;  Conservative   43;  Mismatches   48;  Indels    2;  Gaps    1;

Qy     290 MKLQHVNRKLERRVIDWYQHLQINKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIF 349
           |::: |   |: :|| |: :| : :| ::|   :  ||::|:||:|::||| || ||:||
Db       1 mrmrrvpnhlqvkvikwfdylwltqkcsdeeravsclpdklkaeiainvhldtlkrveif 60

Qy     350 QNCEASLLEELVLKLQPQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGA 409
           || ||  | ||||:|:|  :|||:|:||||::|:||||:  |:| ||||:| |  | | |
Db      61 qnteagflcelvlrlrpvlfspgdyicrkgevgkemyivnrgrlqvvadngktvmaslka 120

Qy     410 GLYFGEISIINIKGNMSGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQAQTIMEEKGR 469
           | |||||||:|:    :||||||:::|:||||| |||:|: :|| ||| |:  :|
Db     121 gsyfgeisilnm--gtagnrrtasvrsvgysdlfvlskkdmwdvlkeypaarvrlesiav 178

Qy     470 EILLKMNK 477
           : | |  |
Db     179 krlekykk 186


RESULT  12
AAB86462
ID   AAB86462 standard; Protein; 1203 AA.
XX
AC   AAB86462;
XX
DT   26-OCT-2001  (first entry)
XX
DE   Human HCN4 protein.
XX
KW   Human; activator; patch-clamp; high through-put screening; inhibitor;
KW   hyperpolarization-activated cation channel; HCN4.
XX
OS   Homo sapiens.
XX
PN   WO200159153-A2.
XX
PD   16-AUG-2001.
XX
PF   24-JAN-2001; 2001WO-EP00755.
XX
PR   12-FEB-2000; 2000DE-1006309.
XX
PA   (AVET ) AVENTIS PHARMA DEUT GMBH.
XX
PI   Jansen H,  Brueggemann A,  Heitsch H,  Goegelein H;
XX
DR   WPI; 2001-550023/61.
DR   N-PSDB; AAH48730.
XX
PT   Identifying substances which modify activity of
PT   hyperpolarization-activated cation channels using fluorescent imaging
PT   is useful to identify possible therapeutic pharmaceuticals  -
XX
PS   Disclosure; Page 39-44; 58pp; German.
XX
CC   This invention describes a novel method for identifying substances which
CC   modify activity of hyperpolarization-activated cation channels (HACC), by
CC   placing cells which express an HACC in an iso-osmotic sodium ion-free
CC   buffer in the presence of a potential-sensitive fluorescent dye, and
CC   measuring membrane potential after the addition of sodium ions. The
CC   invention is used to identify activators or inhibitors of HACCs. Unlike
```

```
CC   prior art patch-clamp techniques, this invention can be automated, and
CC   allows high through put screening. This sequence represents the human
CC   hyperpolarization-activated cation channel protein HCN4 which is
CC   described in the method of the invention.
XX
SQ   Sequence   1203 AA;


  Query Match             15.6%;  Score 467.5;  DB 22;  Length 1203;
  Best Local Similarity   25.1%;  Pred. No. 2.7e-35;
  Matches  125;  Conservative  108;  Mismatches  204;  Indels   61;  Gaps   13;

Qy      26 VLDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHGYLVAWLVLDYTSDLLYLLDMV 85
           :: |  |: ::|  ||:  :: ||||:     |   :      |:| :  ||  :|:|:|
Db     254 iihpysdfrfywdltmlllmvgnliiipvgitf--fkdenttpwivfnvvsdtfflidlv 311

Qy      86 VRFHTGFL--EQGILVVDKGRISSRYVRTWSFFLDLASLMPTDVVY-------------- 129
           : | || :  :   :::|  ||  :|:::| | :|  | :| | ::
Db     312 lnfrtgivvednteiildpqrikmkylksw-fmvdfissipvdyiflivetridsevykt 370

Qy     130 ---VRLGPHTPTLRLNRFLRAPRLFEAFDRTE-----TRTAYPNAFRIAKLMLYIFVVIH 181
             :|:   |  | | | ||  ||     : |     |       ||  |:  : :: |
Db     371 aralrivrftkilsllrllrlsrliryihqweeifhmtydlasavvrivnligmmlilch 430

Qy     182 WNSCLYFALSRYLGFGRDAWVYPDPAQPGFERL-----RRQYLYSFY--FSTLILTTVGD 234
           |: || | :     |  | ||          :     :|| |: :   | ::    |
Db     431 wdgclqflvpmlqdfpddcwv-------sinnmvnnswgkqysyalfkamshmlcigygr 483

Qy     235 TPPPAREEEYLFMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQH 294
             |    : :| |:   ::    :|  :|  :::| :::::   :   :  |::||
Db     484 qapvgmsdvwltml-smivgatcyamfighataliqsldssrrqyqekykqveqymsfhk 542

Qy     295 VNRKLERRVIDWYQHLQINKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEA 354
           :     :|: |:|:| :   || :| :||  | | || |:        :: : :| | :
Db     543 lppdtrqrihdyyeh-ryqgkmfdeesilgelseplreeiinfncrklvasmplfanadp 601

Qy     355 SLLEELVLKLQPQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFG 414
           : :  :: ||: : : ||:|: |:| ||::|| |: | ::|:      :   |  | |||
Db     602 nfvtsmltklrfevfqpgdyiiregtigkkmyfiqhgvvsvltkg--nketkladgsyfg 659

Qy     415 EISIINIKGNMSGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQAQTIMEE-------- 466
           || :      ::  ||||:::  |  |: || ::  ||| |||  :   |
Db     660 eicl------ltrgrrtasvradtycrlyslsvdnfnevleeypmmrrafetvaldrldr 713

Qy     467 --KGREILLKMNKLDVNA 482
             |   |||   : |:|:
Db     714 igkknsillhkvqhdlns 731


RESULT  13
ABB11953
ID   ABB11953 standard; peptide; 1245 AA.
XX
AC   ABB11953;
XX
DT   11-JAN-2002  (first entry)
XX
DE   Human cation channel hHCN4 homologue, SEQ ID NO:2323.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW   haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW   inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW   proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW   myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW   chronic inflammatory condition; proliferative retinopathy;
KW   atherosclerosis; coronary heart disease; arterial ischaemia;
KW   bone disorder; osteoporosis; vascular growth disorder;
KW   tissue regeneration; wound healing; infection; immune disorder;
KW   cell culture; drug screening; gene therapy; antiinflammatory;
KW   antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW   cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW   antifungal; vulnerary; antiulcer.
XX
OS   Homo sapiens.
XX
PN   WO200157188-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US03800.
XX
PR   03-FEB-2000; 2000US-0496914.
PR   27-APR-2000; 2000US-0560875.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-457740/49.
DR   N-PSDB; ABA09197.
XX
PT   Human proteins and DNA encoding sequences useful for preventing,
PT   treating or ameliorating a medical condition in a mammalian subject
PT   e.g. arthritis and cancer -
XX
PS   Claim 20; Page 284; 1963pp; English.
XX
CC   Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
CC   sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC   invention also relates to vectors and recombinant host cells comprising a
CC   nucleotide of the invention, methods of producing the novel polypeptides,
CC   antibodies against the polypeptides, methods of detecting the nucleotides
CC   or polypeptides in a sample, and methods of identifying compounds which
CC   bind to polypeptides of the invention. Although novel, many of the
CC   polypeptides of the invention have homology to known proteins, thereby
CC   giving an insight into their probable biological activities, and hence
CC   potential therapeutic applications. The polypeptides of the invention may
CC   have various activities, including cytokine, cell proliferation or cell
CC   differentiation activities; stem cell growth factor activity;
CC   haematopoiesis regulatory activity; tissue growth activity;
CC   immunomodulatory activity; activin- or inhibin-related activities;
CC   chemotactic or chemokinetic activities; haemostatic, thrombotic or
CC   thrombolytic activities; receptor or ligand activities; or may be
CC   involved in oncogenesis, cancer cell proliferation or metastasis.
CC   Depending on their biological activities, polypeptides and nucleotides of
CC   the invention are useful for preventing, treating or ameliorating medical
CC   conditions, e.g., by protein or gene therapy. Such conditions include
CC   cancers, haematopoietic disorders (e.g.,myeloid or lymphoid cell
CC   disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
CC   proliferative retinopathy, atherosclerosis, coronary heart disease,
CC   arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC   vascular growth. Polypeptides involved with tissue regeneration and
CC   repair (or nucleic acids encoding them) may be used to promote wound
CC   healing (e.g., of burns, incisions and ulcers), while those with
CC   immunomodulatory activities may be used in the treatment of viral,
CC   bacterial and fungal infections in addition to immune disorders.
CC   Polypeptides with growth factor activity may be used in cell cultures to
CC   promote cell growth. For example, such polypeptides may be used to
CC   manipulate stem cells in culture to give rise to neuroepithelial cells
CC   that can be used to augment or replace cells damaged by illness,
CC   autoimmune disease or accidental damage. The polypeptides and nucleotides
CC   may also be used in the diagnosis of the above conditions, and in drug
CC   screening techniques. The present sequence represents a novel human
CC   polypeptide of the invention.
XX
SQ   Sequence   1245 AA;


  Query Match             15.6%;  Score 467.5;  DB 22;  Length 1245;
  Best Local Similarity   25.1%;  Pred. No. 2.8e-35;
  Matches  125;  Conservative  108;  Mismatches  204;  Indels   61;  Gaps   13;

Qy      26 VLDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHGYLVAWLVLDYTSDLLYLLDMV 85
           :: |  |: ::|  ||:  :: ||||:     |   :      |:| :  ||  :|:|:|
Db     296 iihpysdfrfywdltmlllmvgnliiipvgitf--fkdenttpwivfnvvsdtfflidlv 353
```

```
Qy      86 VRFHTGFL--EQGILVVDKGRISSRYVRTWSFFLDLASLMPTDVVY---------- 129
           : | || :  :   :::|  ||  :|:::| | :|  | :| | ::
Db     354 lnfrtgivvednteiildpqrikmkylksw-fmvdfissipvdyiflivetridsevykt 412

Qy     130 ---VRLGPHTPTLRLNRFLRAPRLFEAFDRTE-----TRTAYPNAFRIAKLMLYIFVVIH 181
             :|:   |  | | | ||  ||     : |     |        ||  |:  : :: |
Db     413 aralrivrftkilsllrllrlsrliryihqweeifhmtydlasavvrivnligmmlllch 472

Qy     182 WNSCLYFALSRYLGFGRDAWVYPDPAQPGFERL-----RRQYLYSFY--FSTLILTTVGD 234
           |: || | :     |  | ||          :      :|| |: :   | ::    |
Db     473 wdgclqflvpmlqdfpddcwv-------sinnmvnnswgkqysyalfkamshmlcigygr 525

Qy     235 TPPPAREEEYLFMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQH 294
             |    : :| |:   ::    :|  :|  :::| ::::   :   :  |::||
Db     526 qapvgmsdvwltml-smivgatcyamfighataliqsldssrrqyqekykqveqymsfhk 584

Qy     295 VNRKLERRVIDWYQHLQINKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEA 354
           :    :|: |:|:| :   || :| :||  | | || |:       :: : :| | :
Db     585 lppdtrqrihdyyeh-ryqgkmfdeesilgelseplreeiinfncrklvasmplfanadp 643

Qy     355 SLLEELVLKLQPQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFG 414
           : :  :: ||: : : ||:|: |:| ||::|| |: | ::|:      :   |  | |||
Db     644 nfvtsmltklrfevfqpgdyiiregtigkkmyfiqhgvvsvltkg--nketkladgsyfg 701

Qy     415 EISIINIKGNMSGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQAQTIMEE-------- 466
           || :      ::  ||||:::  |  |: || ::  ||| |||  :   |
Db     702 eicl------ltrgrrtasvradtycrlyslsvdnfnevleeypmmrrafetvaldrldr 755

Qy     467 --KGREILLKMNKLDVNA 482
             |   |||   : |:|:
Db     756 igkknsillhkvqhdlns 773
```

RESULT 14
AAY22193
ID AAY22193 standard; Protein; 528 AA.
XX
AC AAY22193;
XX
DT 10-SEP-1999 (first entry)
XX
DE Mouse brain CNG-2 protein sequence.
XX
KW BCNG; brain cyclic nucleotide gated ion channel; epilepsy; hyperalgesia;
KW Alzheimer's Disease; Parkinson's Disease; long QT syndrome; dyslexia;
KW sick sinus syndrome; age-related memory loss; cystic fibrosis;
KW sudden death syndrome; pacemaker rhythm dysfunction; sensory disorder;
KW auditory disorder; respiratory disorder; attention deficit disorder;
KW learning disability; drug addiction; therapy; mBCNG-2.
XX
OS Mus sp.
XX
PN WO9932615-A1.
XX
PD 01-JUL-1999.
XX
PF 23-DEC-1998; 98WO-US27630.
XX
PR 28-MAY-1998; 98US-0086436.
PR 23-DEC-1997; 97US-0997685.
XX
PA (UYCO) UNIV COLUMBIA NEW YORK.
XX
PI Bartsch D, Grant S, Kandel ER, Santoro B, Siegelbaum S;
PI Tibbs G;
XX
DR WPI; 1999-418922/35.
DR N-PSDB; AAX84444.
XX
PT An isolated nucleic acid encoding a brain or heart cyclic
PT nucleotide-gated ion channel
XX
PS Claim 16; Page 193-195; 213pp; English.
XX
CC This sequence is the brain cyclic nucleotide-gated ion channel
CC (BCNG) of the invention, designated mBCNG-2. BCNG and BCNG-related
CC proteins are useful in screening for compounds that modulate, interact or
CC affect expression. Compounds, e.g. antagonists and agonists, identified
CC in the methods are useful for modulating BCNG or BCNG-related protein
CC activity. Modulation is increased or decreased ion permissivity or ion
CC flow rate. Modulators of BCNG can be used to treat a neurological, renal,
CC pulmonary, hepatic or cardiovascular condition. Such conditions include
CC epilepsy, Alzheimer's Disease, Parkinson's Disease, long QT syndrome,
CC sick sinus syndrome, age-related memory loss, cystic fibrosis, sudden
CC death syndrome or pacemaker rhythm dysfunction. BCNG or BCNG-related
CC protein can also be used to treat sensory disorders, e.g. blindness, loss
CC of vision, loss of smell, numbness and lack of ability to taste. Also
CC treatable are auditory disorders, respiratory disorders due to defects in
CC central nervous system areas that control respiration or defects in the
CC lungs, dyslexia, attention deficit disorder or learning disabilities,
CC drug addiction and regulation of cell secretions. The proteins are useful
CC targets for screening for drugs that are effective in the control of pain
CC and hyperalgesia.
XX
SQ Sequence 528 AA;

Query Match 15.6%; Score 466.5; DB 20; Length 528;
Best Local Similarity 25.3%; Pred. No. 9.2e-36;
Matches 122; Conservative 109; Mismatches 202; Indels 49; Gaps 12;

```
Qy      26 VLDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHGYLVAWLVLDYTSDLLYLLDMV 85
           :: |  |: ::|  ||:  :: ||||:     |   :      |:| :  ||  :|:|:|
Db      27 iihpysdfrfywdftmllfmvgnliipvgitf--fkdettapwivfnvvsdtfflmdlv 84

Qy      86 VRFHTGFL--EQGILVVDKGRISSRYVRTWSFFLDLASLMPTDVVY---------- 129
           : | || :  :   :::|  :|  :|:||| | :|  | :| | ::
Db      85 lnfrtgiviednteiildpekikkkylrtw-fvvdfvssipvdyiflivekgidsevykt 143

Qy     130 ---VRLGPHTPTLRLNRFLRAPRLFEAFDRTE-----TRTAYPNAFRIAKLMLYIFVVIH 181
             :|: | |  | | | ||  ||     : |     |        ||  |:  : :: |
Db     144 aralrivpftkilsllrllrlsrliryihqweeifhmtydlasavmricnlismmlllch 203

Qy     182 WNSCLYFALSRYLGFGRDAWVYPDPAQPGFERLRRQY--LYSF----YFSTLILTTVGDT 235
           |: || | :     |  | ||  :        :   :  ||||      | ::    |
Db     204 wdgclqflvpmlqdfpsdcwvsin------nmvnhswselysfalfkamshmlcigygrq 257

Qy     236 PPPAREEEYLFMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHV 295
           | :  : :| |:   ::    :|  :|  :::| :::::   :   :  |::||    :
Db     258 apesmtdiwltml-smivgatcyamfighataliqsldssrrqyqekykqveqymsfhkl 316

Qy     296 NRKLERRVIDWYQHLQINKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEAS 355
             ::: |:|:| :   ||::| :||  |   || |:        :: : :| | : :
Db     317 padfrqkihdyyeh-ryqgkmsdedsilgelngplreeivnfncrklvasmplfanadpn 375

Qy     356 LLEELVLKLQPQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGE 415
           :  :: ||: : : ||:|: |:| |+::|| |: | ::|:      :   |  | ||||
Db     376 fvtamltklkfevfqpgdyiiregtigkkmyfiqhgvvsvltkg--nkemklsdgsyfge 433

Qy     416 ISIINIKGNMSGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKM 475
           | :      ::  ||||:::  |  |: || ::  ||| |||  :   |    : | ::
Db     434 icl------ltrgrrtasvradtycrlyslsvdnfnevleeypmmrrafetvaidrldri 487

Qy     476 NK 477
           ||
Db     488 gk 489
```

RESULT 15
AAB86461
ID AAB86461 standard; Protein; 889 AA.
XX
AC AAB86461;
XX
DT 26-OCT-2001 (first entry)
XX

DE Human HCN2 protein.
XX
KW Human; activator; patch-clamp; high through-put screening; inhibitor;
KW hyperpolarization-activated cation channel; HCN2.
XX
OS Homo sapiens.
XX
PN WO200159153-A2.
XX
PD 16-AUG-2001.
XX
PF 24-JAN-2001; 2001WO-EP00755.
XX
PR 12-FEB-2000; 2000DE-1006309.
XX
PA (AVET) AVENTIS PHARMA DEUT GMBH.
XX
PI Jansen H, Brueggemann A, Heitsch H, Goegelein H;
XX
DR WPI; 2001-550023/61.
DR N-PSDB; AAH48729.
XX
PT Identifying substances which modify activity of
PT hyperpolarization-activated cation channels using fluorescent imaging
PT is useful to identify possible therapeutic pharmaceuticals -
XX
PS Disclosure; Page 34-37; 58pp; German.
XX
CC This invention describes a novel method for identifying substances which
CC modify activity of hyperpolarization-activated cation channels (HACC), by
CC placing cells which express an HACC in an iso-osmotic sodium ion-free
CC buffer in the presence of a potential-sensitive fluorescent dye, and
CC measuring membrane potential after the addition of sodium ions. The
CC invention is used to identify activators or inhibitors of HACCs. Unlike
CC prior art patch-clamp techniques, this invention can be automated, and
CC allows high through put screening. This sequence represents the human
CC hyperpolarization-activated cation channel protein HCN2 which is
CC described in the method of the invention.
XX
SQ Sequence 889 AA;

Query Match 15.6%; Score 465.5; DB 22; Length 889;
Best Local Similarity 25.5%; Pred. No. 2.6e-35;
Matches 126; Conservative 107; Mismatches 207; Indels 55; Gaps 12;

```
Qy     26 VLDPSGDYYYWWLNTMVFPVMYNLIILVCRACFPDLQHGYLVAWLVLDYTSDLLYLLDMV 85
          :: |  |: ::|  ||:  :: ||||:     |   :      |:| :  ||  :|:|:|
Db    203 iihpysdfrfywdftmllfmvgnliiipvgitf--fkdettapwivfnvvsdtfflmdlv 260

Qy     86 VRFHTGFL--EQGILVVDKGRISSRYVRTWSFFLDLASLMPTDVVY-------------- 129
          : | || :  :   :::|  :|  :|:||| | :|  | :| | ::
Db    261 lnfrtgiviednteiildpekikkkylrtw-fvvdfvssipvdyiflivekgidsevykt 319

Qy    130 ---VRLGPHTPTLRLNRFLRAPRLFEAFDRTE-----TRTAYPNAFRIAKLMLYIFVVIH 181
             :|:   |  | | | ||  ||     : |     |        ||  |:  : :: |
Db    320 aralrivrftkilsllrllrlsrliryihqweeifhmtydlasavmricnlismmlllch 379

Qy    182 WNSCLYFALSRYLGFGRDAWVYPDPAQPGFERLRRQYLYSF----YFSTLILTTVGDTPP 237
          |: || | :     | |: ||    :  |        ||||      | ::    |   |
Db    380 wdgclqflvpmlqdfprncwv----singmvnhswselysfalfkamshmlcigygrqap 435

Qy    238 PAREEEYLFMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNR 297
           :  : :| |:   ::    :|  :|  :::| ::::   :   :  |::||    :
Db    436 esmtdiwltml-smivgatcyamfighataliqsldssrrqyqekykqveqymsfhklpa 494

Qy    298 KLERRVIDWYQHLQINKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLL 357
             ::: |:|:| :   || :| :||  |   || |:       :: : :| | : : :
Db    495 dfrqkihdyyeh-ryqgkmfdedsilgelngplreeivnfncrklvasmplfanadpnfv 553

Qy    358 EELVLKLQPQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEIS 417
           :: ||: : : ||:|: |:| ||::|| |: | ::|:      :   |  | |||||
Db    554 tamltklkfevfqpgdyiiregtigkkmyfiqhgvvsvltkg--nkemklsdgsyfgeic 611

Qy    418 IINIKGNMSGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQAQTIMEE----------K 467
          :      ::  ||||::::  |  |: || ::  ||| |||  :   |           |
Db    612 l------ltrgrrtasvradtycrlyslsvdnfnevleeypmmrrafetvaidrldrigk 665

Qy    468 GREILLKMNKLDVNA 482
             |||   : |:|:
Db    666 knsillhkvqhdlns 680
```

Search completed: June 20, 2002, 17:06:09
Job time: 9442 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 20, 2002, 14:36:07 ; Search time 32.21 Seconds
(without alignments)
436.036 Million cell updates/sec

Title: US-09-927-267-1
Perfect score: 2989
Sequence: 1 MSQDTKVKTTESSPPAPSKA.....EGTSKDEEGRASQEGPPGPE 575

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	380	12.7	1083	4	US-09-600-776-2	Sequence 2, Appli
2	375	12.5	1159	2	US-08-956-242-13	Sequence 13, Appl
3	375	12.5	1159	3	US-09-351-215-13	Sequence 13, Appl
4	375	12.5	1159	4	US-09-226-012-2	Sequence 2, Appli
5	375	12.5	1159	4	US-09-226-012-4	Sequence 4, Appli
6	370.5	12.4	888	2	US-08-956-242-4	Sequence 4, Appli
7	370.5	12.4	888	3	US-09-351-215-4	Sequence 4, Appli
8	361	12.1	1017	4	US-09-600-776-6	Sequence 6, Appli
9	225.5	7.5	626	2	US-08-956-242-2	Sequence 2, Appli
10	225.5	7.5	626	3	US-09-351-215-2	Sequence 2, Appli
11	109	3.6	422	2	US-09-067-351-3	Sequence 3, Appli
12	109	3.6	422	4	US-09-360-490-3	Sequence 3, Appli
13	105.5	3.5	645	2	US-08-592-126-144	Sequence 144, App
14	105.5	3.5	645	2	US-08-687-080-47	Sequence 47, Appl
15	105.5	3.5	1312	2	US-08-592-126-148	Sequence 148, App
16	105.5	3.5	1312	2	US-08-687-080-51	Sequence 51, Appl
17	105	3.5	2101	1	US-08-466-390-4	Sequence 4, Appli
18	105	3.5	2101	1	US-08-470-950-4	Sequence 4, Appli
19	105	3.5	2101	1	US-08-467-781-4	Sequence 4, Appli
20	105	3.5	2101	1	US-08-195-487-4	Sequence 4, Appli
21	105	3.5	2101	2	US-08-483-924-4	Sequence 4, Appli
22	105	3.5	2101	4	US-09-452-294-1	Sequence 1, Appli
23	105	3.5	2101	5	PCT-US93-06160-4	Sequence 4, Appli
24	104	3.5	2547	3	US-09-058-489-35	Sequence 35, Appl
25	103.5	3.5	1886	4	US-08-938-105-3	Sequence 3, Appli
26	102.5	3.4	2482	1	US-08-328-254-6	Sequence 6, Appli
27	101.5	3.4	907	1	US-08-349-006-2	Sequence 2, Appli
28	101.5	3.4	907	3	US-08-804-439A-19	Sequence 19, Appl
29	101.5	3.4	907	3	US-08-720-229-19	Sequence 19, Appl
30	101.5	3.4	907	5	PCT-US94-04180-2	Sequence 2, Appli
31	100.5	3.4	3248	1	US-08-353-700-1	Sequence 1, Appli
32	100.5	3.4	3248	5	PCT-US95-16216-1	Sequence 1, Appli
33	98.5	3.3	553	2	US-08-943-087-2	Sequence 2, Appli
34	98.5	3.3	553	2	US-08-943-087-14	Sequence 14, Appl
35	98.5	3.3	553	2	US-08-943-087-16	Sequence 16, Appl
36	98.5	3.3	553	2	US-08-943-087-18	Sequence 18, Appl
37	98.5	3.3	553	2	US-08-943-087-20	Sequence 20, Appl
38	98.5	3.3	553	2	US-08-943-087-22	Sequence 22, Appl
39	98.5	3.3	553	2	US-08-943-087-24	Sequence 24, Appl
40	98.5	3.3	553	2	US-08-943-087-26	Sequence 26, Appl
41	98.5	3.3	553	2	US-08-943-087-28	Sequence 28, Appl
42	98.5	3.3	553	2	US-08-943-087-30	Sequence 30, Appl
43	98.5	3.3	553	2	US-08-943-087-32	Sequence 32, Appl
44	98.5	3.3	553	2	US-08-943-087-34	Sequence 34, Appl
45	98.5	3.3	553	2	US-08-943-087-36	Sequence 36, Appl

ALIGNMENTS

RESULT 1
US-09-600-776-2
; Sequence 2, Application US/09600776
; Patent No. 6326168
; GENERAL INFORMATION:
; APPLICANT: Yamanouchi Pharmaceutical Co., Ltd.
; TITLE OF INVENTION: A novel potassium channel protein
; FILE REFERENCE: Y9903-PCT
; CURRENT APPLICATION NUMBER: US/09/600,776
; CURRENT FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: JP P1998-011434
; PRIOR FILING DATE: 1998-01-23
; PRIOR APPLICATION NUMBER: JP P1998-346198
; PRIOR FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1083
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-600-776-2

Query Match 12.7%; Score 380; DB 4; Length 1083;
Best Local Similarity 23.5%; Pred. No. 3.1e-29;
Matches 148; Conservative 103; Mismatches 220; Indels 160; Gaps 25;

```
Qy      71 VLDYTSDLLYLLDMVVRFHTGFL-EQGILVVDKGRISSRYVRTWSFFLDLASLMPTDVVY 129
           | |   ::|::||:|: | | |: : | :|     |   || || | ||: : :| |:::
Db     263 VCDLAVEVLFILDIVLNFRTTFVSKSGQVVFAPKSICLHYVTTW-FLLDVIAALPFDLLH 321

Qy     130 -----VRLGPHTPTLRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVI-HWN 183
                |  | |   |:  | ||  ||    ||   :|      :  |:: :| :: ||
Db     322 AFKVNVYFGAH--LLKTVRLLRLLRLLPRLDRYSQYSAV-----VLTLLMAVFALLAHWV 374

Qy     184 SCLYFALSRYLG----------FGRDAWV----------YPDPAQP------------ 209
           :|::|    |:|              |:          |    :|
Db     375 ACVWF----YIGQREIESSESELPEIGWLQELARRLETPYYLVGRRPAGGNSSGQSDNCS 430

Qy     210 ------------GFERLRRQYLYSFYFSTLILTTVGDTPPPAR-EEEYLFMVGDFLLAV 255
                      |   ||  |: | ||:   ||:||     |  : | :| :   |:
Db     431 SSSEANGTGLELLGGPSLRSAYITSLYFALSSLTSVGFGNVSANTDTEKIFSICTMLIGA 490

Qy     256 MGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVIDWYQHLQINKK 315
           :  | : |:::::|  |      ::      :: |:::  : : |::|:::::|
Db     491 LMHAVVFGNVTAIIQRMYARRFLYHSRTRDLRDYIRIHRIPKPLKQRMLEYFQATWAVNN 550

Qy     316 MTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQPQTYSPGEYV 375
             :  :|| ||: |||::|: :|   | :: :|:      |  | | |:|   :||||:
Db     551 GIDTTELLQSLPDELRADIAMHLHKEVL-QLPLFEAASRGCLRALSLALRPAFCTPGEYL 609
```

```
Qy    376 CRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFG-------EISIINIKGNMSGN 428
Db    610 IHQGDALQALYFVCSGSMEVL--KGGTVLAILGKGDLIGCELPRREQV----VKAN---- 659

Qy    429 RRTANIKSLGYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLDVNAEAAEIA 488
Db    660 ---ADVKGLTYCVLQCLQLAGLHDSLALYPEFAPRF-SRGLRGELSYN-LGAGGGSAEV- 713

Qy    489 LQEATESRLRGLDQQLDDLQTK-------------------------------FARLLA 516
Db    714 ----DTSSLSGDNTLMSTLEEKETDGEQGPTVSPAPADEPSSPLLSPGCTSSSSAAKLLS 769

Qy    517 ELESS--------------ALKI--------AYRIERLEWQTREWPMPEDLAEADDEGEP 554
Db    770 PRRTAPRPRLGGRGRPGRAGALKAEAGPSAPPRALEGLRLPPMPWNVPPDLSPRVVDG-I 828

Qy    555 EEGTSKDE---EGRASQEGP-------PGPE 575
Db    829 EDGCGSDQPKFSFRVGQSGPECSSSPSPGPE 859


RESULT   2
US-08-956-242-13
; Sequence 13, Application US/08956242C
; Patent No. 5986081
; GENERAL INFORMATION:
;  APPLICANT: Ganetzky, Barry S.
;  APPLICANT: Titus, Steven A.
;  TITLE OF INVENTION: Polynucleotides Encoding Herg-3
;  FILE REFERENCE: 960296.94550
;  CURRENT APPLICATION NUMBER: US/08/956,242C
;  CURRENT FILING DATE: 1997-10-22
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
;   LENGTH: 1159
;   TYPE: PRT
;   ORGANISM: Homo sapien
US-08-956-242-13


  Query Match             12.5%;  Score 375;  DB 2;  Length 1159;
  Best Local Similarity   22.0%;  Pred. No. 1.1e-28;
  Matches  128;  Conservative  116;  Mismatches  244;  Indels   94;  Gaps   21;

Qy     36 W-WLNTMVFPVMYNLIILVCRACF---------PDLQHGYLVAWL-VLDYTSDLLYLLDM 84
Db    410 WDWL--ILLLVIYTAVFTPYSAAFLLKETEEGPPATECGYACQPLAVVDLIVDIMFIVDI 467

Qy     85 VVRFHTGFLEQGILVVD-KGRISSRYVRTWSFFLDLASLMPTDVVYVRLGPH-----TPT 138
Db    468 LINFRTTYVNANEEVVSHPGRIAVHYFKGW-FLIDMVAAIPFDLLIFGSGSEELIGLLKT 526

Qy    139 LRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSCLYFALSR----YL 194
Db    527 ARLLRLVRVARKLDRYSEY-------GAAVLFLLMCTFALIAHWLACIWYAIGNMEQPHM 579

Qy    195 GFGRDAWVYP---------DPAQPGFERLRRQYLYSFYFSTLILTTVG-DTPPPAREEEY 244
Db    580 D-SRIGWLHNLGDQIGKPYNSSGLGGPSIKDKYVTALYFTFSSLTSVGFGNVSPNTNSEK 638

Qy    245 LFMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVI 304
Db    639 IFSICVMLIGSLMYASIFGNVSAIIQRLYSGTARYHTQMLRVREFIRFHQIPNPLRQRLE 698

Qy    305 DWYQHLQINKKMTNEV---AILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELV 361
Db    699 EYFQHAW---SYTNGIDMNAVLKGFPECLQADICLHLNRSLLQHCKPFRGATKGCLRALA 755

Qy    362 LKLQPQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGE-ISIIN 420
Db    756 MKFKTTHAPPGDTLVHAGDLLTALYFISRGSIEILRGDVVV--AILGKNDIFGEPLNLYA 813

Qy    421 IKGNMSGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLDV 480
Db    814 RPGKSNG-----DVRALTYCDLHKIHRDDLLEVLDMYPE---FSDHFWSSLEITFNLRDT 865

Qy    481 NAEAAEIALQEATESRLRGLDQQLDDLQTKFARLLAELESSALKIAYRIERLEWQTREWP 540
Db    866 NMIPGSPGSTE--------------LEGGFSR------------QRKRKLSFRRRTDK 897

Qy    541 MPEDLAEADDEGEPEEGTSKDEEGR-------ASQEGPPGPE 575
Db    898 DTEQPGEVSALGPGRAGAGPSSRGRPGGPWGESPSSGPSSPE 939


RESULT   3
US-09-351-215-13
; Sequence 13, Application US/09351215
; Patent No. 6087488
; GENERAL INFORMATION:
;  APPLICANT: Ganetzky, Barry S.
;  APPLICANT: Titus, Steven A.
;  TITLE OF INVENTION: Polynucleotides Encoding Herg-3
;  FILE REFERENCE: 960296.94550
;  CURRENT APPLICATION NUMBER: US/09/351,215
;  CURRENT FILING DATE: 1999-07-12
;  EARLIER APPLICATION NUMBER: 08/956,242
;  EARLIER FILING DATE: 1997-10-22
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
;   LENGTH: 1159
;   TYPE: PRT
;   ORGANISM: Homo sapien
US-09-351-215-13


  Query Match             12.5%;  Score 375;  DB 3;  Length 1159;
  Best Local Similarity   22.0%;  Pred. No. 1.1e-28;
  Matches  128;  Conservative  116;  Mismatches  244;  Indels   94;  Gaps   21;

Qy     36 W-WLNTMVFPVMYNLIILVCRACF---------PDLQHGYLVAWL-VLDYTSDLLYLLDM 84
Db    410 WDWL--ILLLVIYTAVFTPYSAAFLLKETEEGPPATECGYACQPLAVVDLIVDIMFIVDI 467

Qy     85 VVRFHTGFLEQGILVVD-KGRISSRYVRTWSFFLDLASLMPTDVVYVRLGPH-----TPT 138
Db    468 LINFRTTYVNANEEVVSHPGRIAVHYFKGW-FLIDMVAAIPFDLLIFGSGSEELIGLLKT 526

Qy    139 LRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSCLYFALSR----YL 194
Db    527 ARLLRLVRVARKLDRYSEY-------GAAVLFLLMCTFALIAHWLACIWYAIGNMEQPHM 579

Qy    195 GFGRDAWVYP---------DPAQPGFERLRRQYLYSFYFSTLILTTVG-DTPPPAREEEY 244
Db    580 D-SRIGWLHNLGDQIGKPYNSSGLGGPSIKDKYVTALYFTFSSLTSVGFGNVSPNTNSEK 638

Qy    245 LFMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVI 304
Db    639 IFSICVMLIGSLMYASIFGNVSAIIQRLYSGTARYHTQMLRVREFIRFHQIPNPLRQRLE 698

Qy    305 DWYQHLQINKKMTNEV---AILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELV 361
Db    699 EYFQHAW---SYTNGIDMNAVLKGFPECLQADICLHLNRSLLQHCKPFRGATKGCLRALA 755

Qy    362 LKLQPQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGE-ISIIN 420
Db    756 MKFKTTHAPPGDTLVHAGDLLTALYFISRGSIEILRGDVVV--AILGKNDIFGEPLNLYA 813

Qy    421 IKGNMSGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLDV 480
Db    814 RPGKSNG-----DVRALTYCDLHKIHRDDLLEVLDMYPE---FSDHFWSSLEITFNLRDT 865

Qy    481 NAEAAEIALQEATESRLRGLDQQLDDLQTKFARLLAELESSALKIAYRIERLEWQTREWP 540
```

```
Db    866 NMIPGSPGSTE--------------LEGGFSR------------QRKRKLSFRRRTDK 897

Qy    541 MPEDLAEADDEGEPEEGTSKDEEGR-------ASQEGPPGPE 575
Db    898 DTEQPGEVSALGPGRAGAGPSSRGRPGGPWGESPSSGPSSPE 939
```

RESULT 4
US-09-226-012-2
; Sequence 2, Application US/09226012
; Patent No. 6207383
; GENERAL INFORMATION:
; APPLICANT: Keating, Mark T.
; APPLICANT: Splawski, Igor
; TITLE OF INVENTION: MUTATIONS IN AND GENOMIC STRUCTURE OF HERG - A LONG QT
; TITLE OF INVENTION: SYNDROME GENE
; FILE REFERENCE: 2323-136
; CURRENT APPLICATION NUMBER: US/09/226,012
; CURRENT FILING DATE: 1999-01-06
; EARLIER APPLICATION NUMBER: 09/122,847
; EARLIER FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1159
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-226-012-2

Query Match 12.5%; Score 375; DB 4; Length 1159;
Best Local Similarity 22.0%; Pred. No. 1.1e-28;
Matches 128; Conservative 116; Mismatches 244; Indels 94; Gaps 21;

```
Qy     36 W-WLNTMVFPVMYNLIILVCRACF---------PDLQHGYLVAWL-VLDYTSDLLYLLDM 84
Db    410 WDWL--ILLLVIYTAVFTPYSAAFLLKETEEGPPATECGYACQPLAVVDLIVDIMFIVDI 467

Qy     85 VVRFHTGFLEQGILVVD-KGRISSRYVRTWSFFLDLASLMPTDVVYVRLGPH-----TPT 138
Db    468 LINFRTTYVNANEEVVSHPGRIAVHYFKGW-FLIDMVAAIPFDLLIFGSGSEELIGLLKT 526

Qy    139 LRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSCLYFALSR----YL 194
Db    527 ARLLRLVRVARKLDRYSEY-------GAAVLFLLMCTFALIAHWLACIWYAIGNMEQPHM 579

Qy    195 GFGRDAWVYP---------DPAQPGFERLRRQYLYSFYFSTLILTTVG-DTPPPAREEEY 244
Db    580 D-SRIGWLHNLGDQIGKPYNSSGLGGPSIKDKYVTALYFTFSSLTSVGFGNVSPNTNSEK 638

Qy    245 LFMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVI 304
Db    639 IFSICVMLIGSLMYASIFGNVSAIIQRLYSGTARYHTQMLRVREFIRFHQIPNPLRQRLE 698

Qy    305 DWYQHLQINKKMTNEV---AILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELV 361
Db    699 EYFQHAW---SYTNGIDMNAVLKGFPECLQADICLHLNRSLLQHCKPFRGATKGCLRALA 755

Qy    362 LKLQPQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGE-ISIIN 420
Db    756 MKFKTTHAPPGDTLVHAGDLLTALYFISRGSIEILRGDVVV--AILGKNDIFGEPLNLYA 813

Qy    421 IKGNMSGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLDV 480
Db    814 RPGKSNG-----DVRALTYCDLHKIHRDDLLEVLDMYPE---FSDHFWSSLEITFNLRDT 865

Qy    481 NAEAAEIALQEATESRLRGLDQQLDDLQTKFARLLAELESSALKIAYRIERLEWQTREWP 540
Db    866 NMIPGSPGSTE--------------LEGGFSR------------QRKRKLSFRRRTDK 897

Qy    541 MPEDLAEADDEGEPEEGTSKDEEGR-------ASQEGPPGPE 575
```


```
Db    898 DTEQPGEVSALGPGRAGAGPSSRGRPGGPWGESPSSGPSSPE 939
```

RESULT 5
US-09-226-012-4
; Sequence 4, Application US/09226012
; Patent No. 6207383
; GENERAL INFORMATION:
; APPLICANT: Keating, Mark T.
; APPLICANT: Splawski, Igor
; TITLE OF INVENTION: MUTATIONS IN AND GENOMIC STRUCTURE OF HERG - A LONG QT
; TITLE OF INVENTION: SYNDROME GENE
; FILE REFERENCE: 2323-136
; CURRENT APPLICATION NUMBER: US/09/226,012
; CURRENT FILING DATE: 1999-01-06
; EARLIER APPLICATION NUMBER: 09/122,847
; EARLIER FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1159
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-226-012-4

Query Match 12.5%; Score 375; DB 4; Length 1159;
Best Local Similarity 22.0%; Pred. No. 1.1e-28;
Matches 128; Conservative 116; Mismatches 244; Indels 94; Gaps 21;

```
Qy     36 W-WLNTMVFPVMYNLIILVCRACF---------PDLQHGYLVAWL-VLDYTSDLLYLLDM 84
Db    410 WDWL--ILLLVIYTAVFTPYSAAFLLKETEEGPPATECGYACQPLAVVDLIVDIMFIVDI 467

Qy     85 VVRFHTGFLEQGILVVD-KGRISSRYVRTWSFFLDLASLMPTDVVYVRLGPH-----TPT 138
Db    468 LINFRTTYVNANEEVVSHPGRIAVHYFKGW-FLIDMVAAIPFDLLIFGSGSEELIGLLKT 526

Qy    139 LRLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVIHWNSCLYFALSR----YL 194
Db    527 ARLLRLVRVARKLDRYSEY-------GAAVLFLLMCTFALIAHWLACIWYAIGNMEQPHM 579

Qy    195 GFGRDAWVYP---------DPAQPGFERLRRQYLYSFYFSTLILTTVG-DTPPPAREEEY 244
Db    580 D-SRIGWLHNLGDQIGKPYNSSGLGGPSIKDKYVTALYFTFSSLTSVGFGNVSPNTNSEK 638

Qy    245 LFMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVI 304
Db    639 IFSICVMLIGSLMYASIFGNVSAIIQRLYSGTARYHTQMLRVREFIRFHQIPNPLRQRLE 698

Qy    305 DWYQHLQINKKMTNEV---AILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELV 361
Db    699 EYFQHAW---SYTNGIDMNAVLKGFPECLQADICLHLNRSLLQHCKPFRGATKGCLRALA 755

Qy    362 LKLQPQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGE-ISIIN 420
Db    756 MKFKTTHAPPGDTLVHAGDLLTALYFISRGSIEILRGDVVV--AILGKNDIFGEPLNLYA 813

Qy    421 IKGNMSGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQAQTIMEEKGREILLKMNKLDV 480
Db    814 RPGKSNG-----DVRALTYCDLHKIHRDDLLEVLDMYPE---FSDHFWSSLEITFNLRDT 865

Qy    481 NAEAAEIALQEATESRLRGLDQQLDDLQTKFARLLAELESSALKIAYRIERLEWQTREWP 540
Db    866 NMIPGSPGSTE--------------LEGGFSR------------QRKRKLSFRRRTDK 897

Qy    541 MPEDLAEADDEGEPEEGTSKDEEGR-------ASQEGPPGPE 575
Db    898 DTEQPGEVSALGPGRAGAGPSSRGRPGGPWGESPSSGPSSPE 939
```


RESULT 6
US-08-956-242-4

; Sequence 4, Application US/08956242C
; Patent No. 5986081
; GENERAL INFORMATION:
; APPLICANT: Ganetzky, Barry S.
; APPLICANT: Titus, Steven A.
; TITLE OF INVENTION: Polynucleotides Encoding Herg-3
; FILE REFERENCE: 960296.94550
; CURRENT APPLICATION NUMBER: US/08/956,242C
; CURRENT FILING DATE: 1997-10-22
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 888
; TYPE: PRT
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (133)
; OTHER INFORMATION: Unidentified at time of filing
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (181)
; OTHER INFORMATION: Unidentified at time of filing
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (438)
; OTHER INFORMATION: Unidentified at time of filing
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (439)
; OTHER INFORMATION: Unidentified at time of filing
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (567)
; OTHER INFORMATION: Unidentified at time of filing
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (568)
; OTHER INFORMATION: Unidentified at time of filing
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (571)
; OTHER INFORMATION: Unidentified at time of filing
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (575)
; OTHER INFORMATION: Unidentified at time of filing
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (580)
; OTHER INFORMATION: Unidentified at time of filing
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (589)
; OTHER INFORMATION: Unidentified at time of filing
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (593)
; OTHER INFORMATION: Unidentified at time of filing
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (599)
; OTHER INFORMATION: Unidentified at time of filing
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (600)
; OTHER INFORMATION: Unidentified at time of filing
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (608)
; OTHER INFORMATION: Unidentified at time of filing
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (610)
; OTHER INFORMATION: Unidentified at time of filing
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (611)
; OTHER INFORMATION: Unidentified at time of filing
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (621)
; OTHER INFORMATION: Unidentified at time of filing
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (624)
; OTHER INFORMATION: Unidentified at time of filing
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (625)
; OTHER INFORMATION: Unidentified at time of filing
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (628)
; OTHER INFORMATION: Unidentified at time of filing
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (629)
; OTHER INFORMATION: Unidentified at time of filing
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (634)
; OTHER INFORMATION: Unidentified at time of filing
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (637)
; OTHER INFORMATION: Unidentified at time of filing
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (646)
; OTHER INFORMATION: Unidentified at time of filing
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (653)
; OTHER INFORMATION: Unidentified at time of filing
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (656)
; OTHER INFORMATION: Unidentified at time of filing
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (660)
; OTHER INFORMATION: Unidentified at time of filing
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (661)
; OTHER INFORMATION: Unidentified at time of filing
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (662)
; OTHER INFORMATION: Unidentified at time of filing
US-08-956-242-4

Query Match 12.4%; Score 370.5; DB 2; Length 888;
Best Local Similarity 22.7%; Pred. No. 2e-28;
Matches 115; Conservative 107; Mismatches 211; Indels 73; Gaps 16;

```
Qy      36 W-WLNTMVFPVMYNLIILVCRACF--------PDLQHGYLVAWL-VLDYTSDLLYLLDMV 85
           | ||  ::  |:|  |     | |          : ||  : | |:|   |::::|::
Db      53 WDWL--ILLLVIYTAIFTPYSAAFLLNDREEQKRRECGYSCSPLNVVDLIVDIMFIIDIL 110

Qy      86 VRFHTGFLEQGILVV-DKGRISSRYVRTWSFFLDLASLMPTDVVYVRLGPHTPT-----L 139
           : | | :: |   || |  :|:  | : | | :|: : :| |::    |    |     |
Db     111 INFRTTYVNQNEEVVSDPAKIAXHYFKGW-FLIDMVAAIPFDLLIFGSGSDETTTLIGLL 169
```

```
Qy     140 RLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVV-IHWNSCLYFALSR----YL 194
           :  | ||  |:    ||     |      :  | : || :  || :|:::|:      ||
Db     170 KTARLLRLVRVXRKLDRYSEYGA-----AVLMLSMCIFALNAHWLACIWYAIGNVERPYL 224

Qy     195 GFGRDAWV----------YPDPAQPGFERLRRQYLYSFYFSTLILTTVG-DTPPPAREEE 243
              :  |:          | |         :: :|: : ||:   ||:||     |    |
Db     225 -TDKIGWLDSLGQQIGKRYNDSDSSSGPSIKDKYVTALYFTFSSLTSVGFGNVSPNTNSE 283

Qy     244 YLFMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRV 303
            :| :   |:  : :|:| |::|::|  : :  | ::     ||:::::   :    | :|:
Db     284 KIFSICVMLIGSLMYASIFGNVSAIIQRLYSGTARYHMQMLRVKEFIRFHQIPNPLRQRL 343

Qy     304 IDWYQHLQINKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLK 363
           :::||         :   :|:  || |:|:: : :: : |   : |:      |  | :|
Db     344 EEYFQHAWTYTNGIDMNMVLKGFPECLQADICLHLNQTLLQNCKAFRGASKGCLRALAMK 403

Qy     364 LQPQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIINIKG 423
           :      |: :   ||:   :| :  | : :     :   |:||    |||:  :  |
Db     404 FKTTHALQGDTLVHCGDVLTALYFLSRGSIEISKXXLVV--AILGKNDIFGEMVHLYAKP 461

Qy     424 NMSGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQ----------------------- 459
             |     |::::| | ||  : :||| |||  ||:
Db     462 GKS----NADVRALTYCDLHKIQREDLLEVLDMYPEFSDHFLTNLELTFNLRHESAKADL 517

Qy     460 --AQTIMEEKGREILLKMNKLDVNAE 483
             :|:: : :|    |:  ||   :|
Db     518 LRSQSMNDSEGDNCKLRRRKLSFESE 543
```

```
RESULT   7
US-09-351-215-4
; Sequence 4, Application US/09351215
; Patent No. 6087488
; GENERAL INFORMATION:
;  APPLICANT: Ganetzky, Barry S.
;  APPLICANT: Titus,. Steven A.
;  TITLE OF INVENTION: Polynucleotides Encoding Herg-3
;  FILE REFERENCE: 960296.94550
;  CURRENT APPLICATION NUMBER: US/09/351,215
;  CURRENT FILING DATE: 1999-07-12
;  EARLIER APPLICATION NUMBER: 08/956,242
;  EARLIER FILING DATE: 1997-10-22
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 888
;   TYPE: PRT
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (133)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (181)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (438)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (439)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (567)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (568)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (571)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (575)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (580)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (589)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (593)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (599)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (600)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (608)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (610)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (611)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (621)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (624)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (625)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (628)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (629)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (634)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (637)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (646)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
```

```
;   LOCATION: (653)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (656)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (660)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (661)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (662)
;   OTHER INFORMATION: Unidentified at time of filing
US-09-351-215-4


  Query Match             12.4%;  Score 370.5;  DB 3;  Length 888;
  Best Local Similarity   22.7%;  Pred. No. 2e-28;
  Matches  115;  Conservative  107;  Mismatches  211;  Indels   73;  Gaps   16;

Qy      36 W-WLNTMVFPVMYNLIILVCRACF--------PDLQHGYLVAWL-VLDYTSDLLYLLDMV 85
           | ||  ::  |:|  |     | |           : ||  : | |:|   |:::::|::
Db      53 WDWL--ILLLVIYTAIFTPYSAAFLLNDREEQKRRECGYSCSPLNVVDLIVDIMFIIDIL 110

Qy      86 VRFHTGFLEQGILVV-DKGRISSRYVRTWSFFLDLASLMPTDVVYVRLGPHTPT-----L 139
           : | | :: |   || |  :|:  | : | | :|: : :| |::    |     |     |
Db     111 INFRTTYVNQNEEVVSDPAKIAXHYFKGW-FLIDMVAAIPFDLLIFGSGSDETTTLIGLL 169

Qy     140 RLNRFLRAPRLFEAFDRTETRTAYPNAFRIAKLMLYIFVV-IHWNSCLYFALSR----YL 194
           :  | ||  |:    ||     |       :  | : || :  || :|:::|:      ||
Db     170 KTARLLRLVRVXRKLDRYSEYGA-----AVLMLSMCIFALNAHWLACIWYAIGNVERPYL 224

Qy     195 GFGRDAWV----------YPDPAQPGFERLRRQYLYSFYFSTLILTTVG-DTPPPAREEE 243
           :  |:           | |         :: :|: : ||:   ||:||     |    |
Db     225 -TDKIGWLDSLGQQIGKRYNDSDSSSGPSIKDKYVTALYFTFSSLTSVGFGNVSPNTNSE 283

Qy     244 YLFMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRV 303
           :| :   |:  : :|:| |::|::|  : :  | ::     ||::::   :   | :|:
Db     284 KIFSICVMLIGSLMYASIFGNVSAIIQRLYSGTARYHMQMLRVKEFIRFHQIPNPLRQRL 343

Qy     304 IDWYQHLQINKKMTNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLK 363
           :::||        :   :|:  || |:|:: : :: : |   : |:      |  | :|
Db     344 EEYFQHAWTYTNGIDMNMVLKGFPECLQADICLHLNQTLLQNCKAFRGASKGCLRALAMK 403

Qy     364 LQPQTYSPGEYVCRKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIINIKG 423
           :      |: :   ||:   :| :  | : :     :   |:||    |||:  :  |
Db     404 FKTTHALQGDTLVHCGDVLTALYFLSRGSIEISKXXLVV--AILGKNDIFGEMVHLYAKP 461

Qy     424 NMSGNRRTANIKSLGYSDLFCLSKEDLREVLSEYPQ----------------------~~ 459
           |     |::::| | ||  : :|||  |||  ||:
Db     462 GKS----NADVRALTYCDLHKIQREDLLEVLDMYPEFSDHFLTNLELTFNLRHESAKADL 517

Qy     460 --AQTIMEEKGREILLKMNKLDVNAE 483
             :|:: : :|    |:  ||   :|
Db     518 LRSQSMNDSEGDNCKLRRRKLSFESE 543


RESULT   8
US-09-600-776-6
; Sequence 6, Application US/09600776
; Patent No. 6326168
; GENERAL INFORMATION:
;  APPLICANT:  Yamanouchi Pharmaceutical Co., Ltd.
;  TITLE OF INVENTION:  A novel potassium channel protein
;  FILE REFERENCE:  Y9903-PCT
;  CURRENT APPLICATION NUMBER: US/09/600,776
;  CURRENT FILING DATE:  2000-07-21
;  PRIOR APPLICATION NUMBER:  JP P1998-011434
;  PRIOR FILING DATE:  1998-01-23
;  PRIOR APPLICATION NUMBER:  JP P1998-346198
;  PRIOR FILING DATE:  1998-12-04
;  NUMBER OF SEQ ID NOS:  12
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 1017
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-600-776-6


  Query Match             12.1%;  Score 361;  DB 4;  Length 1017;
  Best Local Similarity   24.6%;  Pred. No. 2.4e-27;
  Matches  111;  Conservative  101;  Mismatches  203;  Indels   36;  Gaps   12;

Qy      37 WLNTMVFPVMYNLIILVCRACFPDLQHGYLVA--WLVLDYTSDLLYLLDMVVRFHTGFLE 94
           |   ::    |  : :    ||       : :   || |   ::|::||::: | | ::
Db     229 WDGLILLATFYVAVTVPYNVCFSGDDDTPITSRHTLVSDIAVEMLFILDIILNFRTTYVS 288

Qy      95 QGILVVDKGR-ISSRYVRTWSFFLDLASLMPTDVVYVRLGPHTPTLRLNRFLRAPRLFEA 153
           |   |:   | |   |: || ||:|| : :| |::|:     |  : | : :|  ||
Db     289 QSGQVISAPRSIGLHYLATW-FFIDLIAALPFDLLYIFNITVTSLVHLLKTVRLLRLLRL 347

Qy     154 FDRTETRTAYPNAFRIAKLMLYIFVVIHWNSCLYFALSRYLGFGRD------AWV----- 202
           : | | :  :|  :  ||    :: || :|::: : |      |       |:
Db     348 LQKLE-RYSQCSAVVLTLLMSVFALLAHWMACIWYVIGRREMEANDPLLWDIGWLHELGK 406

Qy     203 -----YPDPAQPGFERLRRQYLYSFYFSTLILTTVGDTPPPAR-EEEYLFMVGDFLLAVM 256
                | : :  |  | |  |: : ||:   ||:||     |  : | :| :   |:  :
Db     407 RLEVPYVNGSVGGPSR-RSAYIAALYFTLSSLTSVGFGNVCANTDAEKIFSICTMLIGAL 465

Qy     257 GFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVIDWYQHLQINKKM 316
             | : |:::::|  | :  : ::     :| ::::  : | |::|:::::|
Db     466 MHAVVFGNVTAIIQRMYSRRSLYHSRMKDLKDFIRVHRLPRPLKQRMLEYFQTTWAVNSG 525

Qy     317 TNEVAILQHLPERLRAEVAVSVHLSTLSRVQIFQNCEASLLEELVLKLQPQTYSPGEYVC 376
           :   :|:  |: |||::|: ::   | :: :|       |  | | ::    :||||:
Db     526 IDANELLRDFPDELRADIAMHLNREIL-QLPLFGAASRGCLRALSLHIKTSFCAPGEYLL 584

Qy     377 RKGDIGQEMYIIREGQLAVVADDGITQYAVLGAGLYFGEISIINIKGNMSG--------N 428
           |:|| |  | :  | | |: |: :   |:|| |   |  : |   |   |
Db     585 RRGDALQAHYYVCSGSLEVLRDNMV--LAILGKGDLIG--ADIPEPGQEPGLGADPNFVL 640

Qy     429 RRTANIKSLGYSDLFCLSKEDLREVLSEYPQ 459
           : :|::|:| |  |  ||   | |||  ||:
Db     641 KTSADVKALTYCGLQQLSSRGLAEVLRLYPE 671


RESULT   9
US-08-956-242-2
; Sequence 2, Application US/08956242C
; Patent No. 5986081
; GENERAL INFORMATION:
;  APPLICANT: Ganetzky, Barry S.
;  APPLICANT: Titus, Steven A.
;  TITLE OF INVENTION: Polynucleotides Encoding Herg-3
;  FILE REFERENCE: 960296.94550
;  CURRENT APPLICATION NUMBER: US/08/956,242C
;  CURRENT FILING DATE: 1997-10-22
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 626
;   TYPE: PRT
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (441)..(542)
;   OTHER INFORMATION: Unidentified at time of filing
US-08-956-242-2
```

```
  Query Match             7.5%;  Score 225.5;  DB 2;  Length 626;
  Best Local Similarity   21.3%;  Pred. No. 6.2e-14;
  Matches   78;  Conservative   74;  Mismatches  130;  Indels   85;  Gaps   13;

Qy      12 SSPPAPSKARKLLPVLDP----------------SGDYYYWW--LNTMVFPVMYN---L 49
           | | :|: |   |  | |                       : |  |:   |  :::   |
Db      78 SRPVSPALAAPCLTSLGPHPQVLSLGADVLPEYKLQAPRIHRWTILHYSPFKAVWDWLIL 137

Qy      50 IILVCRACFPDLQHGYLVA----------------WLVLDYTSDLLYLLDMVVRFHTGFL 93
           ::::  | |      :|::                 |:|   |:::::|:|: | | ::
Db     138 LLVIYTAVFTPYSAAFLLSDQDESRRGACSYTCSPLTVVDLIVDIMFVVDIVINFRTTYV 197

Qy      94 EQGILVVD-KGRISSRYVRTWSFFLDLASLMPTDVVYVRLGPHTPT-----LRLNRFLRA 147
             ||    ||:  | : | | :|: : :| |::  | |    |     |:  | ||
Db     198 NTNDEVVSHPRRIAVHYFKGW-FLIDMVAAIPFDLLIFRTGSDETTTLIGLLKTARLLRL 256

Qy     148 PRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVI-HWNSCLYFAL---------------- 190
           |:    ||     |      :  |::  | :| || :|:::|:
Db     257 VRVARKLDRYSEYGA-----AVLFLLMCTFALIAHWLACIWYAIGNVERPYLEHKIGWLD 311

Qy     191 -------SRYLGFGRDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVG-DTPPPAREE 242
              || |         |||      :: :|: : ||:   ||:||     |
Db     312 SLAVQLGKRYNG--------SDPASG--PSVQDKYVTALYFTFSSLTSVGFGNVSPNTNS 361

Qy     243 EYLFMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERR 302
           | :| :   |:  : :|:| |::|::|  :  :  | ::     ||:::   :   | :|
Db     362 EKVFSICVMLIGSLMYASIFGNVSAIIQRLYSGTARYHMQMLRVKEFIRFHQIPNPLRQR 421

Qy     303 VIDWYQH 309
           : :::||
Db     422 LEEYFQH 428


RESULT  10
US-09-351-215-2
; Sequence 2, Application US/09351215
; Patent No. 6087488
; GENERAL INFORMATION:
;  APPLICANT: Ganetzky, Barry S.
;  APPLICANT: Titus, Steven A.
;  TITLE OF INVENTION: Polynucleotides Encoding Herg-3
;  FILE REFERENCE: 960296.94550
;  CURRENT APPLICATION NUMBER: US/09/351,215
;  CURRENT FILING DATE: 1999-07-12
;  EARLIER APPLICATION NUMBER: 08/956,242
;  EARLIER FILING DATE: 1997-10-22
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 626
;   TYPE: PRT
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (441)..(542)
;   OTHER INFORMATION: Unidentified at time of filing
US-09-351-215-2


  Query Match             7.5%;  Score 225.5;  DB 3;  Length 626;
  Best Local Similarity   21.3%;  Pred. No. 6.2e-14;
  Matches   78;  Conservative   74;  Mismatches  130;  Indels   85;  Gaps   13;

Qy      12 SSPPAPSKARKLLPVLDP----------------SGDYYYWW--LNTMVFPVMYN---L 49
           | | :|: |   |  | |                       : |  |:   |  :::   |
Db      78 SRPVSPALAAPCLTSLGPHPQVLSLGADVLPEYKLQAPRIHRWTILHYSPFKAVWDWLIL 137

Qy      50 IILVCRACFPDLQHGYLVA----------------WLVLDYTSDLLYLLDMVVRFHTGFL 93
           ::::  | |      :|::                 |:|   |:::::|:|: | | ::
Db     138 LLVIYTAVFTPYSAAFLLSDQDESRRGACSYTCSPLTVVDLIVDIMFVVDIVINFRTTYV 197

Qy      94 EQGILVVD-KGRISSRYVRTWSFFLDLASLMPTDVVYVRLGPHTPT-----LRLNRFLRA 147
             ||    ||:  | : | | :|: : :| |::  | |    |     |:  | ||
Db     198 NTNDEVVSHPRRIAVHYFKGW-FLIDMVAAIPFDLLIFRTGSDETTTLIGLLKTARLLRL 256

Qy     148 PRLFEAFDRTETRTAYPNAFRIAKLMLYIFVVI-HWNSCLYFAL---------------- 190
           |:    ||     |      :  |::  | :| || :|:::|:
Db     257 VRVARKLDRYSEYGA-----AVLFLLMCTFALIAHWLACIWYAIGNVERPYLEHKIGWLD 311

Qy     191 -------SRYLGFGRDAWVYPDPAQPGFERLRRQYLYSFYFSTLILTTVG-DTPPPAREE 242
              || |         |||      :: :|: : ||:   ||:||     |
Db     312 SLAVQLGKRYNG--------SDPASG--PSVQDKYVTALYFTFSSLTSVGFGNVSPNTNS 361

Qy     243 EYLFMVGDFLLAVMGFATIMGSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERR 302
           | :| :   |:  : :|:| |::|::|  :  :  | ::     ||:::   :   | :|
Db     362 EKVFSICVMLIGSLMYASIFGNVSAIIQRLYSGTARYHMQMLRVKEFIRFHQIPNPLRQR 421

Qy     303 VIDWYQH 309
           : :::||
Db     422 LEEYFQH 428


RESULT  11
US-09-067-351-3
; Sequence 3, Application US/09067351
; Patent No. 5994081
;  GENERAL INFORMATION:
;    APPLICANT:  Tang, Y. Tom
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Corley, Neil C.
;    APPLICANT:  Baughn, Mariah
;    TITLE OF INVENTION:  HUMAN KERATINS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  INCYTE PHARMACEUTICALS, INC.
;      STREET:  3174 PORTER DRIVE
;      CITY:  PALO ALTO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Word Perfect 6.1 for Windows/MS-DOS 6.2
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/067,351
;      FILING DATE:  Herewith
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CERRONE, MICHAEL C.
;      REGISTRATION NUMBER:  39,132
;      REFERENCE/DOCKET NUMBER:  PF-0511 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 855-0555
;      TELEFAX:  (650) 845-4166
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  422 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  BRSTNOT07
;      CLONE:  2124178
US-09-067-351-3


  Query Match             3.6%;  Score 109;  DB 2;  Length 422;
  Best Local Similarity   20.2%;  Pred. No. 0.022;
  Matches   74;  Conservative   55;  Mismatches  121;  Indels  116;  Gaps   16;

Qy     263 GSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVIDWYQ---------- 308
           |:  : : |:|   |::      ::|   |:  | |||  |:: |:|
Db      70 GNGKATMQNLNDRLASY------LEKVRALEEANMKLESRILKWHQQRDPGSKKDYSQYE 123
```

```
Qy    309 ----HLQ---INKKMTNEVAILQHLPERLR------------------AEVAVSVHLST 342
              |||   :: ||||   ||  | :  |                    |: |     |
Db    124 ENITHLQEQIVDGKMTNAQIIL--LIDNARMAVDDFNLKYENEHSFKKDLEIEVEGLRRT 181

Qy    343 LSRVQIF-----QNCEASLLEELVLKLQPQT-----YSPGEY-VCRKGDIGQEMYIIREG 391
          |  : |       |  |    | :::|   :      : | :: |  | | |    :|:
Db    182 LDNLTIVTTDLEQEVEGMRKELILMKKHHEQEMEKHHVPSDFNVNVKVDTGPREDLIK-- 239

Qy    392 QLAVVADDGITQYAVLGAGLYFGEISIINIKGNMSGNRRTANIKSLGYSDLFCLSKEDLR 451
            | :|   :| :           ||  |               : ||    ||
Db    240 ----VLEDMRQEYEL----------IIKKK---------------HRDLDTWYKEQSA 268

Qy    452 EVLSEYPQAQTIMEEKGREILLKMNKLDVNAEAAEIALQEATESRLRGLDQQLDDLQTKF 511
            :  |     |:   :|      :::|    :| || ||    ::   |:  | : |:::
Db    269 AMSQEAASPATVQSRQG-----DIHELKRTFQALEIDLQTQYSTK-SALENMLSETQSRY 322

Qy    512 A-------RLLAELESSALKIAYRIERL--EWQT---------REWPMPEDLAEADDEGE 553
            :        :::  |    :: : :||   |:|         :|      | | : ||
Db    323 SCKLQDMQEIISHYEEELTQLRHELERQNNEYQVLLGIKTHLEKEITTYRRLLEGESEGT 382

Qy    554 PEEGTS 559
            ||  |
Db    383 REESKS 388
```

RESULT 12
US-09-360-490-3
; Sequence 3, Application US/09360490
; Patent No. 6221843
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Baughn, Mariah
; TITLE OF INVENTION: HUMAN KERATINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/360,490
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/067,351
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: CERRONE, MICHAEL C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0511 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 422 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRSTNOT07
; CLONE: 2124178
US-09-360-490-3

Query Match 3.6%; Score 109; DB 4; Length 422;
Best Local Similarity 20.2%; Pred. No. 0.022;
Matches 74; Conservative 55; Mismatches 121; Indels 116; Gaps 16;

```
Qy    263 GSMSSVIYNMNTADAAFYPDHALVKKYMKLQHVNRKLERRVIDWYQ------------- 308
          |:  : : |:|   |::      ::|   |:  | ||| |:: |:|
Db     70 GNGKATMQNLNDRLASY------LEKVRALEEANMKLESRILKWHQQRDPGSKKDYSQYE 123

Qy    309 ----HLQ---INKKMTNEVAILQHLPERLR------------------AEVAVSVHLST 342
              |||   :: ||||   ||  | :  |                    |: |     |
Db    124 ENITHLQEQIVDGKMTNAQIIL--LIDNARMAVDDFNLKYENEHSFKKDLEIEVEGLRRT 181

Qy    343 LSRVQIF-----QNCEASLLEELVLKLQPQT-----YSPGEY-VCRKGDIGQEMYIIREG 391
          |  : |       |  |    | :::|   :      : | :: |  | | |    :|:
Db    182 LDNLTIVTTDLEQEVEGMRKELILMKKHHEQEMEKHHVPSDFNVNVKVDTGPREDLIK-- 239

Qy    392 QLAVVADDGITQYAVLGAGLYFGEISIINIKGNMSGNRRTANIKSLGYSDLFCLSKEDLR 451
            | :|   :| :           ||  |               : ||    ||
Db    240 ----VLEDMRQEYEL----------IIKKK---------------HRDLDTWYKEQSA 268

Qy    452 EVLSEYPQAQTIMEEKGREILLKMNKLDVNAEAAEIALQEATESRLRGLDQQLDDLQTKF 511
            :  |     |:   :|      :::|    :| || ||    ::   |:  | : |:::
Db    269 AMSQEAASPATVQSRQG-----DIHELKRTFQALEIDLQTQYSTK-SALENMLSETQSRY 322

Qy    512 A-------RLLAELESSALKIAYRIERL--EWQT---------REWPMPEDLAEADDEGE 553
            :        :::  |    :: : :||   |:|         :|      | | : ||
Db    323 SCKLQDMQEIISHYEEELTQLRHELERQNNEYQVLLGIKTHLEKEITTYRRLLEGESEGT 382

Qy    554 PEEGTS 559
            ||  |
Db    383 REESKS 388
```

RESULT 13
US-08-592-126-144
; Sequence 144, Application US/08592126
; Patent No. 5821091
; GENERAL INFORMATION:
; APPLICANT: Gregory Dolganov
; TITLE OF INVENTION: Transcripts Encoding Immunomodulatory
; TITLE OF INVENTION: Polypeptides
; NUMBER OF SEQUENCES: 151
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,126
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; REFERENCE/DOCKET NUMBER: 4600-0111
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 144:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 645 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein

```
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  G18.pep
US-08-592-126-144


  Query Match             3.5%;  Score 105.5;  DB 2;  Length 645;
  Best Local Similarity   20.0%;  Pred. No. 0.1;
  Matches   61;  Conservative   52;  Mismatches  119;  Indels   73;  Gaps    9;

Qy     282 DHALVKKYMKLQHVNRKLERRVIDWYQHLQINKKMTNEVAILQH-----LPERLRAEVAV 336
           |  |   |   |   |: ||:::| :: |:   |:  |  :|       | |: | ::
Db     301 DEQLNDLYHNHQRTVREKERKLVDCHRELE---KLNKESRLLNQEKSELLVEQGRLQLQA 357

Qy     337 SVHLSTLSRVQIFQNCEASLLEELVLKLQPQTYSPGEYVCRKGDIGQEMYIIREGQLAVV 396
             |       |       ||:: |  :|:   :  | :  |:  |     ::|| |
Db     358 DRH-------QEHIRARDSLIQSLATQLELDGFERGPFSERQ--IKNFHKLVRERQ---- 404

Qy     397 ADDGITQYAVLGAGLYFGEISIINIKGNMSGNRRTANIKSLGYSDLFCLSKEDLREVLSE 456
                                     |  :|||     :::   | :: : |:  :
Db     405 ----------------------------EGEAKTANQLMNDFAEKETLKQKQIDEIRDK 435

Qy     457 YPQAQTIMEEKGREILLKMNKL-DVNAEAAEIALQEATESRLRGLDQQLDDLQTKFARLL 515
             |:| |   :  | |:| :|  |  ::   | :  |:  |||:|     |  | |
Db     436 KTGLGRIIELKSEILSKKQNELKNVKYELQQL---EGSSDRILELDQEL----IKAEREL 488

Qy     516 AELESSALKIAYRIERLEWQTREWPMPEDLAEADDEGE--------------PEEGTSK 560
           :: | ::     ::| :  |  :  :   | : | | |                ::   |
Db     489 SKAEKNSNVETLKMEVISLQNEKADLDRTLRKLDQEMEQLNHHTTTRTQMEMLTKDKADK 548

Qy     561 DEEGR 565
           ||: |
Db     549 DEQIR 553


RESULT  14
US-08-687-080-47
; Sequence 47, Application US/08687080
; Patent No. 5965427
;  GENERAL INFORMATION:
;    APPLICANT:  Gregory Dolganov
;    TITLE OF INVENTION:  Human RAD50 Gene and Methods of Use Thereof
;    NUMBER OF SEQUENCES:  175
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dehlinger & Associates
;      STREET:  350 Cambridge Avenue, Suite 250
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/687,080
;      FILING DATE:  17-JUL-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/592,126
;      FILING DATE:  26-JAN-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sholtz, Charles K.
;      REGISTRATION NUMBER:  38,615
;      REFERENCE/DOCKET NUMBER:  4600-0111.30
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 324-0880
;      TELEFAX:  (415) 324-0960
;  INFORMATION FOR SEQ ID NO:  47:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  645 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  G18.pep
US-08-687-080-47


  Query Match             3.5%;  Score 105.5;  DB 2;  Length 645;
  Best Local Similarity   20.0%;  Pred. No. 0.1;
  Matches   61;  Conservative   52;  Mismatches  119;  Indels   73;  Gaps    9;

Qy     282 DHALVKKYMKLQHVNRKLERRVIDWYQHLQINKKMTNEVAILQH-----LPERLRAEVAV 336
           |  |   |   |   |: ||:::| :: |:   |:  |  :|       | |: | ::
Db     301 DEQLNDLYHNHQRTVREKERKLVDCHRELE---KLNKESRLLNQEKSELLVEQGRLQLQA 357

Qy     337 SVHLSTLSRVQIFQNCEASLLEELVLKLQPQTYSPGEYVCRKGDIGQEMYIIREGQLAVV 396
             |       |       ||:: |  :|:   :  | :  |:  |     ::|| |
Db     358 DRH-------QEHIRARDSLIQSLATQLELDGFERGPFSERQ--IKNFHKLVRERQ---- 404

Qy     397 ADDGITQYAVLGAGLYFGEISIINIKGNMSGNRRTANIKSLGYSDLFCLSKEDLREVLSE 456
                                     |  :|||     :::   | :: : |:  :
Db     405 ----------------------------EGEAKTANQLMNDFAEKETLKQKQIDEIRDK 435

Qy     457 YPQAQTIMEEKGREILLKMNKL-DVNAEAAEIALQEATESRLRGLDQQLDDLQTKFARLL 515
             |:| |   :  | |:| :|  |  ::   | :  |:  |||:|     |  | |
Db     436 KTGLGRIIELKSEILSKKQNELKNVKYELQQL---EGSSDRILELDQEL----IKAEREL 488

Qy     516 AELESSALKIAYRIERLEWQTREWPMPEDLAEADDEGE--------------PEEGTSK 560
           :: | ::     ::| :  |  :  :   | : | | |                ::   |
Db     489 SKAEKNSNVETLKMEVISLQNEKADLDRTLRKLDQEMEQLNHHTTTRTQMEMLTKDKADK 548

Qy     561 DEEGR 565
           ||: |
Db     549 DEQIR 553


RESULT  15
US-08-592-126-148
; Sequence 148, Application US/08592126
; Patent No. 5821091
;  GENERAL INFORMATION:
;    APPLICANT:  Gregory Dolganov
;    TITLE OF INVENTION:  Transcripts Encoding Immunomodulatory
;    TITLE OF INVENTION:  Polypeptides
;    NUMBER OF SEQUENCES:  151
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dehlinger & Associates
;      STREET:  350 Cambridge Avenue, Suite 250
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/592,126
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sholtz, Charles K.
;      REGISTRATION NUMBER:  38,615
;      REFERENCE/DOCKET NUMBER:  4600-0111
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 324-0880
;      TELEFAX:  (415) 324-0960
;  INFORMATION FOR SEQ ID NO:  148:
```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1312 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Rad50.pro-translation of SEQ ID NO:54
US-08-592-126-148

Query Match 3.5%; Score 105.5; DB 2; Length 1312;
Best Local Similarity 20.0%; Pred. No. 0.34;
Matches 61; Conservative 52; Mismatches 119; Indels 73; Gaps 9;

```
Qy    282 DHALVKKYMKLQHVNRKLERRVIDWYQHLQINKKMTNEVAILQH-----LPERLRAEVAV 336
          |  |  .|   |   |: ||:::| :: |:   |:  |  :|       | |: | ::
Db    301 DEQLNDLYHNHQRTVREKERKLVDCHRELE---KLNKESRLLNQEKSELLVEQGRLQLQA 357

Qy    337 SVHLSTLSRVQIFQNCEASLLEELVLKLQPQTYSPGEYVCRKGDIGQEMYIIREGQLAVV 396
            |       |       ||:: |  :|:   :  | :  |:  |     ::|| |
Db    358 DRH-------QEHIRARDSLIQSLATQLELDGFERGPFSERQ--IKNFHKLVRERQ---- 404

Qy    397 ADDGITQYAVLGAGLYFGEISIIINIKGNMSGNRRTANIKSLGYSDLFCLSKEDLREVLSE 456
                                     |  :|||     :::   | :: : |:  :
Db    405 ----------------------------EGEAKTANQLMNDFAEKETLKQKQIDEIRDK 435

Qy    457 YPQAQTIMEEKGREILLKMNKL-DVNAEAAEIALQEATESRLRGLDQQLDDLQTKFARLL 515
                |:| |   :  | |:| :|  |  ::   | :  |:  |||:|     |  | |
Db    436 KTGLGRIIELKSEILSKKQNELKNVKYELQQL---EGSSDRILELDQEL----IKAEREL 488

Qy    516 AELESSALKIAYRIERLEWQTREWPMPEDLAEADDEGE--------------PEEGTSK 560
          :: | ::     ::| :  |  :  :   | : | | |                ::   |
Db    489 SKAEKNSNVETLKMEVISLQNEKADLDRTLRKLDQEMEQLNHHTTTRTQMEMLTKDKADK 548

Qy    561 DEEGR 565
          ||: |
Db    549 DEQIR 553
```

Search completed: June 20, 2002, 17:06:55
Job time: 9048 sec

OM of: US-09-927-267-1 to: Issued_Patents_NA:* out_format : pfs

Date: Jun 20, 2002 6:30 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09927267/runat_19062002_090143_25504/app_query.fasta_1.640
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09927267_@CGN1_1_94 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-927-267-1
Query length: 575
Database: Issued_Patents_NA:*
Database sequences: 383533
Database length: 122816752
Search time (sec): 79.760000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-600-776-9	+	382.50	755.67	2.1e-34	3715	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-600-776-1	+	380.00	751.85	3.4e-34	3323	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-226-012-1	+	375.00	740.66	1.4e-33	3480	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-226-012-3	+	375.00	739.00	1.8e-33	3950	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-956-242-3	+	370.50	732.48	4.1e-33	3142	!
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-351-215-3	+	370.50	732.48	4.1e-33	3142	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-600-776-10	+	362.00	712.22	5.5e-32	3736	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-600-776-5	+	361.00	712.71	5.2e-32	3064	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-172-422-3	+	352.50	705.91	1.2e-31	1307	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-956-242-1	+	231.00	437.29	1.1e-16	3141	!
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-351-215-1	+	231.00	437.29	1.1e-16	3141	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-795-788-1	+	219.50	417.80	1.4e-15	2173	!
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US93-03942-1	+	219.50	417.80	1.4e-15	2173	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-2	+	122.50	112.51	140.50	4403765	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-067-351-6	+	109.00	187.80	0.0090	1623	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-360-490-6	+	109.00	187.80	0.0090	1623	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-592-126-54	+	105.50	163.46	0.2042	5893	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-687-080-44	+	105.50	163.46	0.2042	5893	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-153-804-3	+	105.00	177.77	0.0325	1828	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-693-457-1	+	105.00	170.67	0.0809	3139	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-265-731-1	+	105.00	170.67	0.0809	3139	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-466-390-3	+	105.00	161.51	0.2622	6306	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-470-950-3	+	105.00	161.51	0.2622	6306	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-467-781-3	+	105.00	161.51	0.2622	6306	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-195-487-3	+	105.00	161.51	0.2622	6306	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-483-924-3	+	105.00	161.51	0.2622	6306	!
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US93-06160-3	+	105.00	161.51	0.2622	6306	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-658-665-42	+	104.50	172.34	0.0653	2550	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-658-665-44	+	104.50	172.34	0.0653	2550	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-796-101-6	+	104.50	172.34	0.0653	2550	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-796-101-8	+	104.50	172.34	0.0653	2550	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-085-273-42	+	104.50	172.34	0.0653	2550	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-085-273-44	+	104.50	172.34	0.0653	2550	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-658-665-37	+	104.50	171.48	0.0730	2724	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-796-101-1	+	104.50	171.48	0.0730	2724	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-085-273-37	+	104.50	171.48	0.0730	2724	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-658-665-38	-	104.50	165.60	0.1551	4260	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-796-101-2	-	104.50	165.60	0.1551	4260	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-085-273-38	-	104.50	165.60	0.1551	4260	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-658-665-43	-	104.50	164.61	0.1761	4594	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-658-665-45	-	104.50	164.61	0.1761	4594	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-796-101-7	-	104.50	164.61	0.1761	4594	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-796-101-9	-	104.50	164.61	0.1761	4594	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-085-273-43	-	104.50	164.61	0.1761	4594	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-085-273-45	-	104.50	164.61	0.1761	4594	

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-600-776-9

seq_documentation_block:
; Sequence 9, Application US/09600776
; Patent No. 6326168
; GENERAL INFORMATION:
; APPLICANT: Yamanouchi Pharmaceutical Co., Ltd.
; TITLE OF INVENTION: A novel potassium channel protein
; FILE REFERENCE: Y9903-PCT
; CURRENT APPLICATION NUMBER: US/09/600,776
; CURRENT FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: JP P1998-011434
; PRIOR FILING DATE: 1998-01-23
; PRIOR APPLICATION NUMBER: JP P1998-346198
; PRIOR FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 3715
; TYPE: DNA
; ORGANISM: Rattus sp.
US-09-600-776-9

alignment_scores:
Quality: 382.50 Length: 624
Ratio: 1.242 Gaps: 22
Percent Similarity: 49.359 Percent Identity: 22.917

alignment_block:
US-09-927-267-1 x US-09-600-776-9 ..

Align seg 1/1 to: US-09-600-776-9 from: 1 to: 3715

```
      71 ValLeuAspTyrThrSerAspLeuLeuTyrLeuLeuAspMetValValAr 87
         |||   |||   :::   ::::::|||::::::||||||:::|||:::::
     942 GTCTGTGACCTGGCCGTGGAAGTCCTCTTCATCTTAGATATTGTGCTGAA 991

      87 gPheHisThrGlyPheLeu...GluGlnGlyIleLeuValValAspLysG 103
         :|||:::|||   |||:::   ::::::::|   :::|||
     992 TTTTCGTACTACATTTGTGTCCAAGTCAGGCCAGGTGGTATTCGCCCCAA 1041

     103 lyArgIleSerSerArgTyrValArgThrTrpSerPhePheLeuAspLeu 119
             |||      :::|||||||   ||||||   |||:::|||||:::
    1042 AGTCCATTTGCCTCCACTACGTCACCACCTGG...TTCCTGCTGGATGTC 1088

     120 AlaSerLeuMetProThrAspValValTyr.............ValAr 131
         :::   :::|||   |||:::::::::                 |||
    1089 ATAGCAGCACTGCCCTTTGACCTACTACATGCCTTCAAGGTCAATGTGTA 1138

     131 gLeuGlyProHisThrProThrLeuArgLeuAsnArgPheLeuArgAlaP 148
         :::|||   |||       |||:::     |||:::||||||
    1139 CGTTGGGGCTCAC......CTACTGAAGACCGTGCGGCTGCTTCGGCTGC 1182

     148 roArgLeuPheGluAlaPheAspArgThrGluThrArgThrAlaTyrPro 164
           ||||||:::      :::||||||   :::      :::|||
    1183 TGCGCCTACTACCAAGACTGGACCGGTACTCTCAGTATAGCGCTGTT... 1229

     165 AsnAlaPheArgIleAlaLysLeuMetLeuTyrIlePheValValIle.. 180
                 :::      |||::::::   :::|||:::::::::
    1230 ............GTGCTCACCTTGCTCATGGCTGTGTTTGCCCTGCTCGC 1267

     181 .HisTrpAsnSerCysLeuTyrPheAlaLeuSerArgTyrLeuGly.... 195
         ||||||   :::|||:::::|||           |||:::|||
    1268 CCACTGGGTGGCCTGCGTTTGGTTC............TACATCGGCCAGC 1305

     196 .........................PheGlyArgAspAlaTrpValTyr 203
```

```
                                    :::   :::   :::|||:::
1306 AAGAGATTGAGAACAGCGAGTCAGAGCTGCCTGAGATCGGCTGGCTGCAG 1355

 204 ..............................................Pr 204
                                                   ||
1356 GAGCTGGCACGCAGGCTGGAGACGCCCTATTACCTGGTGAGCCGGAGTCC 1405

 204 oAspProAlaGlnPro.................................. 209
     ||||   ::::::
1406 AGATGGAGGGAACAGCTCTGGCCAGAGTGAAAACTGCAGTAGCAGTGGCG 1455

 210 .....................................GlyPheGluArg 213
                                           |||
1456 GCGGCAGCGAAGCCAACGGGACTGGGCTGGAGCTGCTGGGTGGCCCATCC 1505

 214 LeuArgArgGlnTyrLeuTyrSerPheTyrPheSerThrLeuIleLeuTh 230
     ||||||     |||:::  |||:::|||||||:::       |||||
1506 CTACGCAGCGCCTACATCACCTCCTTGTACTTCGCGCTCAGCAGTCTCAC 1555

 230 rThrValGlyAspThrProProProAlaArg...GluGluGluTyrLeuP 246
     |:::||||||              |||:::   :::   |||   :::|
1556 CAGTGTGGGCTTCGGCAATGTGTCCGCTAACACAGACACTGAGAAGATTT 1605

 246 heMetValGlyAspPheLeuLeuAlaValMetGlyPheAlaThrIleMet 262
     ||   :::     :::|||:::::::::::     |||:::::::::
1606 TCTCCATCTGCACCATGCTTATTGGAGCTCTGATGCATGCAGTGGTGTTT 1655

 263 GlySerMetSerSerValIleTyrAsnMetAsnThrAlaAspAlaAlaPh 279
     |||:::::::::::::::|||   :::|||   :::          ::
1656 GGGAATGTGACAGCCATCATCCAGCGCATGTACGCTCGCCGCTTTCTGTA 1705

 279 eTyrProAspHisAlaLeuValLysLysTyrMetLysLeuGlnHisValA 296
     ::::               ::::::   |||:::::::::::::::::::
1706 CCACAGCCGCACCCGTGACCTGCGAGACTACATTCGCATCCACCGCATCC 1755

 296 snArgLysLeuGluArgArgValIleAspTrpTyrGlnHisLeuGlnIle 312
       :::   |||:::::::|||:::::::::::::::|||
1756 CCAAGCCCCTCAAGCAGCGCATGCTCGAGTACTTCCAAGCCACCTGGGCC 1805

 313 AsnLysLysMetThrAsnGluValAlaIleLeuGlnHisLeuProGluAr 329
       ::::::      :::  :::   :::||||||   ||||||:::::
1806 GTGAACAACGGCATCGATACCACTGAGCTGCTGCAGAGCCTTCCGGATGA 1855

 329 gLeuArgAlaGluValAlaValSerValHisLeuSerThrLeuSerArgV 346
     :|||||||||:::::::|||:::   :::|||   :::::::|||   ::::
1856 GCTTCGAGCAGACATCGCCATGCACCTGCACAAGGAGGTCCTG...CAGC 1902

 346 alGlnIlePheGlnAsnCysGluAlaSerLeuLeuGluGluLeuValLeu 362
     ::   :::|||:::   ::::::   :::   |||:::   |||   |||
1903 TGCCATTGTTCGAGGCAGCGAGCCGTGGCTGCCTGCGGGCACTGTCTCTG 1952

 363 LysLeuGlnProGlnThrTyrSerProGlyGluTyrValCysArgLysGl 379
       |||:::|||         :::|||||||||||||:::   ::::::||
1953 GCCCTGAGGCCCGCCTTCTGCACGCCGGGCGAGTACCTCATTCACCAAGG 2002

 379 yAspIleGlyGlnGluMetTyrIleIleArgGluGlyGlnLeuAlaValV 396
     ||||      |||   :::|||::::::   :::|||::::::   |||:
2003 CGATGCTCTCCAGGCTCTCTACTTTGTGTGCTCAGGTTCCATGGAGGTCC 2052

 396 alAlaAspAspGlyIleThrGlnTyrAlaValLeuGlyAlaGlyLeuTyr 412
     ::         |||   |||      |||:::::::   |||
2053 TC......AAAGGTGGCACCGTCCTCGCCATTCTAGGGAAGGGTGACCTG 2096

 413 PheGly...................GluIleSerIleIleAsnIleLy 422
     :::|||                 ::::::           :::||
2097 ATCGGCTGTGAGCTGCCCCAGCGAGAGCAAGTA...........GTGAA 2134

 422 sGlyAsnMetSerGlyAsnArgArgThrAlaAsnIleLysSerLeuGlyT 439
     |:::|||                    |||::::::|||:::|||   |
2135 GGCCAAT...................GCCGACGTAAAGGGGCTGACAT 2163

 439 yrSerAspLeuPheCysLeuSerLysGluAspLeuArgGluValLeuSer 455
     ||      |||   ||||||:::         |||:::|||   |||:::

     2164 ACTGCGTCCTACAGTGCCTGCAGCTGGCTGGGCTGCACGAGAGCCTCGCA 2213

      456 GluTyrProGlnAlaGlnThrIleMetGluGluLysGlyArgGluIleLe 472
          ||||||:::           :::::::::      |||
     2214 CTGTACCCTGAGTTTGCCCCACGCTTTAGCCGTGGCCTCCGA........ 2255

      472 uLeuLysMetAsnLysLeuAspValAsnAlaGluAlaAlaGluIleAlaL 489
                ::::::|||:::   |||      |||:::   ::::::
     2256 ..........GGGGAGCTCAGCTACAACCTGGGAGCTGGAGGAGTGTCTG 2295

      489 euGlnGluAlaThrGluSerArgLeuArgGlyLeuAspGlnGlnLeuAsp 505
          :::      |||:::|||   |||   |||   :::      ::::::
     2296 CAGAGGTGGATACCAGCTCA...CTGAGTGGTGACAACACCCTCATGTCC 2342

      506 AspLeuGlnThrLys..................................... 510
          |||:::   |||
     2343 ACACTGGAGGAGAAGGAGACAGATGGGGAGCAAGGACACACGATCTCACC 2392

      510 ..................................................... 510

     2393 AGCCCCAGCAGATGAGCCCTCCAGCCCCCTGCTGTCACCTGGCTGTACCT 2442

      511 ...........PheAlaArgLeuLeuAlaGlu..................... 517
                     |||:::||||||:::
     2443 CCTCCTCCTCAGCGGCCAAACTACTCTCCCCACGTCGAACTGCACCCCGG 2492

      518 ..........................................LeuGluSe 520
                                                   |||:::
     2493 CCGAGGCTGGGTGGCAGAGGGCGGCCAAGTAGGGCAGGGGTTTTGAAGCC 2542

      520 rSerAlaLeuLysIleAlaTyr......ArgIleGluArgLeuGluTrpG 535
            :::|||         |||:::       ::::::   |||:::
     2543 TGAGGCTGGTCCTTCTGCTCATCCACGGACACTTGATGGGTTGCAGCTGC 2592

      535 lnThrArgGluTrpProMetProGluAspLeuAlaGluAlaAspAspGlu 551
                |||   :::|||   ||||||:::           :::
     2593 CCCCCATGCCATGGAATGTACCTCCAGACCTGAGCCCCAGGGTCGTAGAT 2642

      552 GlyGluProGluGluGlyThrSerLysAspGluGluGlyArgAlaSerGl 568
          |||      |||:::|||   ::::::|||::::::      :::   ::
     2643 GGC...ATTGAGGATGGCTGCGGCTCTGACCAGCACAAGTTCTCTTTCCG 2689

      568 nGluGlyProProGlyProGlu 575
          :   |||      |||||||||
     2690 GGTGGGTCAGTCTGGCCCAGAA 2711
```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-600-776-1

```
seq_documentation_block:
; Sequence 1, Application US/09600776
; Patent No. 6326168
; GENERAL INFORMATION:
;  APPLICANT:  Yamanouchi Pharmaceutical Co., Ltd.
;  TITLE OF INVENTION:  A novel potassium channel protein
;  FILE REFERENCE:  Y9903-PCT
;  CURRENT APPLICATION NUMBER: US/09/600,776
;  CURRENT FILING DATE:  2000-07-21
;  PRIOR APPLICATION NUMBER:  JP P1998-011434
;  PRIOR FILING DATE:  1998-01-23
;  PRIOR APPLICATION NUMBER:  JP P1998-346198
;  PRIOR FILING DATE:  1998-12-04
;  NUMBER OF SEQ ID NOS:  12
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 3323
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (6)..(3257)
US-09-600-776-1
```

```
alignment_scores:
            Quality:  380.00              Length:    631
              Ratio:  1.218                Gaps:     25
 Percent Similarity:  49.445   Percent Identity: 23.455

alignment_block:
 US-09-927-267-1 x US-09-600-776-1   ..

 Align seg 1/1  to: US-09-600-776-1  from: 1  to: 3323

      71 ValLeuAspTyrThrSerAspLeuLeuTyrLeuLeuAspMetValValAr 87
         |||   |||   :::   :::::::|||:::::::||||||:::|||:::::
     792 GTCTGTGACCTGGCCGTGGAGGTCCTCTTCATCCTTGACATTGTGCTGAA 841

      87 gPheHisThrGlyPheLeu...GluGlnGlyIleLeuValValAspLysG 103
         :|||:::|||   |||:::   :::::::|||   :::|||
     842 TTTCCGTACCACATTCGTGTCCAAGTCGGGCCAGGTGGTGTTTGCCCCAA 891

     103 lyArgIleSerSerArgTyrValArgThrTrpSerPhePheLeuAspLeu 119
           |||       :::|||||||   ||||||   |||:::||||||:::
     892 AGTCCATTTGCCTCCACTACGTCACCACCTGG...TTCCTGCTGGATGTC 938

     120 AlaSerLeuMetProThrAspValValTyr..............ValAr 131
         :::   :::|||   |||::::::::::                |||
     939 ATCGCAGCGCTGCCCTTTGACCTGCTACATGCCTTCAAGGTCAACGTGTA 988

     131 gLeuGlyProHisThrProThrLeuArgLeuAsnArgPheLeuArgAlaP 148
          :::|||   |||         |||:::      |||:::||||||
     989 CTTCGGGGCCCAT......CTGCTGAAGACGGTGCGCCTGCTGCGCCTGC 1032

     148 roArgLeuPheGluAlaPheAspArgThrGluThrArgThrAlaTyrPro 164
           ||||||:::      :::||||||   :::      :::|||
    1033 TGCGCCTGCTTCCGCGGCTGGACCGGTACTCGCAGTACAGCGCCGTG... 1079

     165 AsnAlaPheArgIleAlaLysLeuMetLeuTyrIlePheValValIle.. 180
                  :::      |||::::::   :::|||::::::::::
    1080 ............GTGCTGACACTGCTCATGGCCGTGTTCGCCCTGCTCGC 1117

     181 .HisTrpAsnSerCysLeuTyrPheAlaLeuSerArgTyrLeuGly.... 195
          ||||||   :::|||:::::::|||         |||:::|||
    1118 GCACTGGGTCGCCTGCGTCTGGTTT............TACATTGGCCAGC 1155

     196 ........................PheGlyArgAspAlaTrpVal... 202
                                 :::   :::   :::|||:::
    1156 GGGAGATCGAGAGCAGCGAATCCGAGCTGCCTGAGATTGGCTGGCTGCAG 1205

     203 ............................TyrProAspProAlaGlnPr 209
                                     |||            :::||
    1206 GAGCTGGCCCGCCGACTGGAGACTCCCTACTACCTGGTGGGCCGGAGGCC 1255

     209 o..................................................... 209
         |
    1256 AGCTGGAGGGAACAGCTCCGGCCAGAGTGACAACTGCAGCAGCAGCAGCG 1305

     210 ..........................GlyPheGluArgLeuArgArg 216
                                  |||          ||||||
    1306 AGGCCAACGGGACGGGGCTGGAGCTGCTGGGCGGCCCGTCGCTGCGCAGC 1355

     217 GlnTyrLeuTyrSerPheTyrPheSerThrLeuIleLeuThrThrValGl 233
         |||:::   |||:::|||||||::         ||||||::||||
    1356 GCCTACATCACCTCCCTCTACTTCGCACTCAGCAGCCTCACCAGCGTGGG 1405

     233 yAspThrProProProAlaArg...GluGluGluTyrLeuPheMetValG 249
         |          |||:::   :::   |||   :::|||   :::
    1406 CTTCGGCAACGTGTCCGCCAACACGGACACCGAGAAGATCTTCTCCATCT 1455

     249 lyAspPheLeuLeuAlaValMetGlyPheAlaThrIleMetGlySerMet 265
           :::|||:::::::::::      |||:::::::::|||::::::
    1456 GCACCATGCTCATCGGCGCCCTGATGCACGCGGTGGTGTTTGGGAACGTG 1505

     266 SerSerValIleTyrAsnMetAsnThrAlaAspAlaAlaPheTyrProAs 282
         :::::::::|||   :::|||   :::            ::::::
    1506 ACGGCCATCATCCAGCGCATGTACGCCCGCCGCTTTCTGTACCACAGCCG 1555

     282 pHisAlaLeuValLysLysTyrMetLysLeuGlnHisValAsnArgLysL 299
                 ::::::   |||:::::::::::::::::::   :::   |
    1556 CACGCGCGACCTGCGCGACTACATCCGCATCCACCGTATCCCCAAGCCCC 1605

     299 euGluArgArgValIleAspTrpTyrGlnHisLeuGlnIleAsnLysLys 315
         ||:::::::|||:::::::::::::::|||                ::::::
    1606 TCAAGCAGCGCATGCTGGAGTACTTCCAGGCCACCTGGGCGGTGAACAAT 1655

     316 MetThrAsnGluValAlaIleLeuGlnHisLeuProGluArgLeuArgAl 332
               :::   :::   :::|||||||   ||||||::::::|||||||
    1656 GGCATCGACACCACCGAGCTGCTGCAGAGCCTCCCTGACGAGCTGCGCGC 1705

     332 aGluValAlaValSerValHisLeuSerThrLeuSerArgValGlnIleP 349
         |::::::|||:::   :::|||   :::::::|||   ::::::   :::|
    1706 AGACATCGCCATGCACCTGCACAAGGAGGTCCTG...CAGCTGCCACTGT 1752

     349 heGlnAsnCysGluAlaSerLeuLeuGluGluLeuValLeuLysLeuGln 365
         ||:::   ::::::   :::   |||:::   |||   |||   |||:::
    1753 TTGAGGCGGCCAGCCGCGGCTGCCTGCGGGCACTGTCTCTGGCCCTGCGG 1802

     366 ProGlnThrTyrSerProGlyGluTyrValCysArgLysGlyAspIleGl 382
         |||         :::|||||||||||||:::   ::::::||||||
    1803 CCCGCCTTCTGCACGCCGGGCGAGTACCTCATCCACCAAGGCGATGCCCT 1852

     382 yGlnGluMetTyrIleIleArgGluGlyGlnLeuAlaValValAlaAspA 399
          |||   :::|||::::::   :::|||::::::   |||:::
    1853 GCAGGCCCTCTACTTTGTCTGCTCTGGCTCCATGGAGGTGCTC......A 1896

     399 spGlyIleThrGlnTyrAlaValLeuGlyAlaGlyLeuTyrPheGly... 414
           |||   |||       |||:::||||||   |||       :::|||
    1897 AGGGTGGCACCGTGCTCGCCATCCTAGGGAAGGGCGACCTGATCGGCTGT 1946

     415 ..................GluIleSerIleIleAsnIleLysGlyAsnMe 425
                            ::::::            :::|||:::|||
    1947 GAGCTGCCCCGGCGGGAGCAGGTG............GTAAAGGCCAAT.. 1982

     425 tSerGlyAsnArgArgThrAlaAsnIleLysSerLeuGlyTyrSerAspL 442
                          |||::::::|||:::|||   |||       |
    1983 ..................GCCGACGTGAAGGGGCTGACGTACTGCGTCC 2013

     442 euPheCysLeuSerLysGluAspLeuArgGluValLeuSerGluTyrPro 458
         ||   ||||||:::          |||::::::   |||:::   ||||||
    2014 TGCAGTGTCTGCAGCTGGCTGGCCTGCACGACAGCCTTGCGCTGTACCCC 2063

     459 GlnAlaGlnThrIleMetGluGluLysGlyArgGluIleLeuLeuLysMe 475
         :::             :::   ::::::|||   :::       |||:::
    2064 GAGTTTGCCCCGCGCTTC...AGTCGTGGCCTCCGAGGGGAGCTCAGCTA 2110

     475 tAsnLysLeuAspValAsnAlaGluAlaAlaGluIleAlaLeuGlnGluA 492
          |||   |||   :::::::::   :::||||||:::
    2111 CAAC...CTGGGTGCTGGGGGAGGCTCTGCAGAGGTG..... 2144

     492 laThrGluSerArgLeuArgGlyLeuAspGlnGlnLeuAspAspLeuGln 508
               |||   |||   |||   :::       ::::::   |||:::
    2145 ..GACACCAGCTCCCTGAGCGGCGACAATACCCTTATGTCCACGCTGGAG 2192

     509 ThrLys..................................... 510
         |||
    2193 GAGAAGGAGACAGATGGGGAGCAGGGCCCCACGGTCTCCCCAGCCCCAGC 2242

     510 ..................................................... 510

    2243 TGATGAGCCCTCCAGCCCCCTGCTGTCCCCTGGCTGCACCTCCTCATCCT 2292

     511 ..PheAlaArgLeuLeuAlaGluLeuGluSerSer.............. 521
           |||:::|||||:::      :::::::::
    2293 CAGCTGCCAAGCTGCTATCCCCACGTCGAACAGCACCCCGGCCTCGTCTA 2342

     522 .....................AlaLeuLysIle........ 525
                              |||||||||
    2343 GGTGGCAGAGGGAGGCCAGGCAGGGCAGGGGCTTTGAAGGCTGAGGCTGG 2392
```

```
     526 .............AlaTyrArgIleGluArgLeuGluTrpGlnThrArgG 538
                          :::|||   |||:::
    2393 CCCCTCTGCTCCCCCACGGGCCCTAGAGGGGCTACGGCTGCCCCCCATGC 2442

     538 luTrpProMetProGluAspLeuAlaGluAlaAspAspGluGlyGluPro 554
           |||   ::::|||   ||||||:::            :::|||
    2443 CATGGAATGTGCCCCCAGATCTGAGCCCCAGGGTAGTAGATGGC...ATT 2489

     555 GluGluGlyThrSerLysAspGlu.........GluGlyArgAlaSerGl 568
         |||:::|||   ::::::|||:::         :::   |||::::::||
    2490 GAAGACGGCTGTGGCTCGGACCAGCCCAAGTTCTCTTTCCGCGTGGGCCA 2539

     568 nGluGlyPro.....................ProGlyProGlu 575
         |:::||||||                   ||||||||||||
    2540 GTCTGGCCCGGAATGTAGCAGCAGCCCCTCCCCTGGACCAGAG 2582

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-226-012-1

seq_documentation_block:
; Sequence 1, Application US/09226012
; Patent No. 6207383
; GENERAL INFORMATION:
;  APPLICANT: Keating, Mark T.
;  APPLICANT: Splawski, Igor
;  TITLE OF INVENTION: MUTATIONS IN AND GENOMIC STRUCTURE OF HERG - A LONG QT
;  TITLE OF INVENTION: SYNDROME GENE
;  FILE REFERENCE: 2323-136
;  CURRENT APPLICATION NUMBER: US/09/226,012
;  CURRENT FILING DATE: 1999-01-06
;  EARLIER APPLICATION NUMBER: 09/122,847
;  EARLIER FILING DATE: 1998-07-27
;  NUMBER OF SEQ ID NOS: 116
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 3480
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(3477)
US-09-226-012-1


alignment_scores:
            Quality:  375.00            Length:    582
              Ratio:  1.154               Gaps:     21
 Percent Similarity:  55.842   Percent Identity: 21.993

alignment_block:
 US-09-927-267-1 x US-09-226-012-1   ..

 Align seg 1/1  to: US-09-226-012-1  from: 1  to: 3480

      36 Trp...TrpLeuAsnThrMetValProProValMetTyrAsnLeuIleIl 51
         |||   ||||||      :::::::::   |||:::|||:::   :::::
    1228 TGGGACTGGCTC......ATCCTGCTGCTGGTCATCTACACGGCTGTCTT 1271

      51 eLeuValCysArgAlaCysPhe.........................P 59
         :           |||:::|||                          |
    1272 CACACCCTACTCGGCTGCCTTCCTGCTGAAGGAGACGGAAGAAGGCCCGC 1321

      59 roAspLeuGlnHisGlyTyrLeuValAlaTrpLeu...ValLeuAspTyr 74
         ||      :::   ||||||          |||   |||:::|||
    1322 CTGCTACCGAGTGTGGCTACGCCTGCCAGCCGCTGGCTGTGGTGGACCTC 1371

      75 ThrSerAspLeuLeuTyrLeuLeuAspMetValValArgPheHisThrGl 91
            |||:::::::::::::::|||:::::::::::::|||:::|||
    1372 ATCGTGGACATCATGTTCATTGTGGACATCCTCATCAACTTCCGCACCAC 1421

      91 yPheLeuGluGlnGlyIleLeuValValAsp...LysGlyArgIleSerS 107
         :::::::::   :::      ||||||:::      |||||||||:::
    1422 CTACGTCAATGCCAACGAGGAGGTGGTCAGCCACCCCGGCCGCATCGCCG 1471

     107 erArgTyrValArgThrTrpSerPhePheLeuAspLeuAlaSerLeuMet 123
           :::|||   :::   |||   |||:::::::|||:::::::::   :::
    1472 TCCACTACTTCAAGGGCTGG...TTCCTCATCGACATGGTGGCCGCCATC 1518

     124 ProThrAspValValTyrValArgLeuGlyProHis.............. 135
         |||   |||::::::            |||   :::
    1519 CCCTTCGACCTGCTCATCTTCGGCTCTGGCTCTGAGGAGCTGATCGGGCT 1568

     136 .ThrProThrLeuArgLeuAsnArgPheLeuArgAlaProArgLeuPheG 152
              |||   ||||||   |||:::::|||:::   |||   ::::
    1569 GCTGAAGACTGCGCGGCTGCTGCGGCTGGTGCGCGTGGCGCGGAAGCTGG 1618

     152 luAlaPheAspArgThrGluThrArgThrAlaTyrProAsnAlaPheArg 168
         ::   :::::::::                        :::|||
    1619 ATCGCTACTCAGAGTAC.....................GGCGCGGCCGTG 1647

     169 IleAlaLysLeuMetLeuTyrIlePheValValIleHisTrpAsnSerCy 185
         :::      ||||||      :::   ::::::   ||||||   :::||
    1648 CTGTTCTTGCTCATGTGCACCTTTGCGCTCATCGCGCACTGGCTAGCCTG 1697

     185 sLeuTyrPheAlaLeuSerArg............TyrLeuGlyPheGlyA 198
         |:::::::::|||:::::::::            ::::::      :::|
    1698 CATCTGGTACGCCATCGGCAACATGGAGCAGCCACACATGGAC...TCAC 1744

     198 rgAspAlaTrpValTyrPro.........................Asp 205
         ||   :::|||::::::                           :::
    1745 GCATCGGCTGGCTGCACAACCTGGGCGACCAGATAGGCAAACCCTACAAC 1794

     206 ProAlaGlnProGlyPheGluArgLeuArgArgGlnTyrLeuTyrSerPh 222
            :::      |||         ::::::   :::|||:::   :::::
    1795 AGCAGCGGCCTGGGCGGCCCCTCCATCAAGGACAAGTATGTGACGGCGCT 1844

     222 eTyrPheSerThrLeuIleLeuThrThrValGly...AspThrProProP 238
         :||||||:::         ||||||:::||||||      :::      |
    1845 CTACTTCACCTTCAGCAGCCTCACCAGTGTGGGCTTCGGCAACGTCTCTC 1894

     238 roAlaArgGluGluGluTyrLeuPheMetValGlyAspPheLeuLeuAla 254
         ||      ::::::|||   :::|||   :::      :::|||::::::
    1895 CCAACACCAACTCAGAGAAGATCTTCTCCATCTGCGTCATGCTCATTGGC 1944

     255 ValMetGlyPheAlaThrIleMetGlySerMetSerSerValIleTyrAs 271
            :::   :::|||:::|||:::|||:::::::|||:::::::|||   ::
    1945 TCCCTCATGTATGCTAGCATCTTCGGCAACGTGTCGGCCATCATCCAGCG 1994

     271 nMetAsnThrAlaAspAlaAlaPheTyrProAspHisAlaLeuValLysL 288
         ::::   ::::::   |||   ::::::   :::          |||:::::
    1995 GCTGTACTCGGGCACAGCCCGCTACCACACACAGATGCTGCGGGTGCGGG 2044

     288 ysTyrMetLysLeuGlnHisValAsnArgLysLeuGluArgArgValIle 304
         ::::::::::::::::::::::::   :::   |||:::::::|||:::
    2045 AGTTCATCCGCTTCCACCAGATCCCCAATCCCCTGCGCCAGCGCCTCGAG 2094

     305 AspTrpTyrGlnHisLeuGlnIleAsnLysLysMetThrAsnGluVal.. 320
         :::::::::||||||               :::   ||||||   :::
    2095 GAGTACTTCCAGCACGCCTGG.........TCCTACACCAACGGCATCGA 2135

     321 .......AlaIleLeuGlnHisLeuProGluArgLeuArgAlaGluValA 335
                |||:::|||:::   :::||||||   |||:::|||:::::::
    2136 CATGAACGCGGTGCTGAAGGGCTTCCCTGAGTGCCTGCAGGCTGACATCT 2185

     335 laValSerValHisLeuSerThrLeuSerArgValGlnIlePheGlnAsn 351
         :::::   ::::::   |||   |||::::::   :::   |||:::::::
    2186 GCCTGCACCTGAACCGCTCACTGCTGCAGCACTGCAAACCCTTCCGAGGG 2235

     352 CysGluAlaSerLeuLeuGluGluLeuValLeuLysLeuGlnProGlnTh 368
         :::      :::   |||:::   |||::::::|||::::::
    2236 GCCACCAAGGGCTGCCTTCGGGCCCTGGCCATGAAGTTCAAGACCACACA 2285

     368 rTyrSerProGlyGluTyrValCysArgLysGlyAspIleGlyGlnGluM 385
                |||||||:::   :::   :::   ||||||:::         :
    2286 TGCACCGCCAGGGGACACACTGGTGCATGCTGGGGACCTGCTCACCGCCC 2335

     385 etTyrIleIleArgGluGlyGlnLeuAlaValValAlaAspAspGlyIle 401
```

```
      ::|||:::|||   :::|||:::::::   ::::::      |||   :::
   2336 TGTACTTCATCTCCCGGGGCTCCATCGAGATCCTGCGGGGCGACGTCGTC 2385

    402 ThrGlnTyrAlaValLeuGlyAlaGlyLeuTyrPheGlyGlu...IleSe 417
      :::      |||:::||||||   :::      |||||||||   :::::
   2386 GTG......GCCATCCTGGGGAAGAATGACATCTTTGGGGAGCCTCTGAA 2429

    417 rIleIleAsnIleLysGlyAsnMetSerGlyAsnArgArgThrAlaAsnI 434
      ::::           |||:::   :::|||              ::::
   2430 CCTGTATGCAAGGCCTGGCAAGTCGAACGGG.............GATG 2464

    434 leLysSerLeuGlyTyrSerAspLeuPheCysLeuSerLysGluAspLeu 450
      :::::::::|||   |||   ||||||      :::   ::::::||||||
   2465 TGCGGGCCCTCACCTACTGTGACCTACACAAGATCCATCGGGACGACCTG 2514

    451 ArgGluValLeuSerGluTyrProGlnAlaGlnThrIleMetGluGluLy 467
        ||||||||||:::   ||||||:::         :::   ::::::
   2515 CTGGAGGTGCTGGACATGTACCCTGAG.........TTCTCCGACCACTT 2555

    467 sGlyArgGluIleLeuLeuLysMetAsnLysLeuAspValAsnAlaGluA 484
          ::::::   :::   :::|||      |||:::|||
   2556 CTGGTCCAGCCTGGAGATCACCTTCAACCTGCGAGATACCAACATGATCC 2605

    484 laAlaGluIleAlaLeuGlnGluAlaThrGluSerArgLeuArgGlyLeu 500
      ::::::   :::      |||
   2606 CGGGCTCCCCCGGCAGTACGGAG........................... 2628

    501 AspGlnGlnLeuAspAspLeuGlnThrLysPheAlaArgLeuLeuAlaGl 517
                        |||:::      |||:::|||
   2629 .................TTAGAGGGTGGCTTCAGTCGG........... 2649

    517 uLeuGluSerSerAlaLeuLysIleAlaTyrArgIleGluArgLeuGluT 534
                                    |||   :::::::|||::::
   2650 ..........................CAACGCAAGCGCAAGTTGTCCT 2671

    534 rpGlnThrArgGluTrpProMetProGluAspLeuAlaGluAlaAspAsp 550
      :::::   |||                |||:::   :::|||::::::
   2672 TCCGCAGGCGCACGGACAAGGACACGGAGCAGCCAGGGGAGGTGTCGGCC 2721

    551 GluGlyGluProGluGluGlyThrSerLysAspGluGluGlyArg..... 565
        |||      :::   |||:::::::   ::::::::::::::::
   2722 TTGGGGCCGGGCCGGGCGGGGGCAGGGCCGAGTAGCCGGGGCCGGCCGGG 2771

    566 ................AlaSerGlnGluGlyProProGlyProGlu 575
                       :::   ::::::||||||   :::||||||
   2772 GGGGCCGTGGGGGGAGAGCCCGTCCAGTGGCCCCTCCAGCCCTGAG 2817

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-226-012-3

seq_documentation_block:
; Sequence 3, Application US/09226012
; Patent No. 6207383
; GENERAL INFORMATION:
;  APPLICANT: Keating, Mark T.
;  APPLICANT: Splawski, Igor
;  TITLE OF INVENTION: MUTATIONS IN AND GENOMIC STRUCTURE OF HERG - A LONG QT
;  TITLE OF INVENTION: SYNDROME GENE
;  FILE REFERENCE: 2323-136
;  CURRENT APPLICATION NUMBER: US/09/226,012
;  CURRENT FILING DATE: 1999-01-06
;  EARLIER APPLICATION NUMBER: 09/122,847
;  EARLIER FILING DATE: 1998-07-27
;  NUMBER OF SEQ ID NOS: 116
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;  LENGTH: 3950
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (67)..(3543)
US-09-226-012-3

alignment_scores:
          Quality:  375.00          Length:    582
            Ratio:  1.154             Gaps:     21
 Percent Similarity:  55.842   Percent Identity: 21.993

alignment_block:
 US-09-927-267-1 x US-09-226-012-3   ..

 Align seg 1/1  to: US-09-226-012-3  from: 1  to: 3950

     36 Trp...TrpLeuAsnThrMetValProValMetTyrAsnLeuIleIl 51
        |||   ||||||      :::::::::   |||:::|||:::   :::::
   1294 TGGGACTGGCTC......ATCCTGCTGCTGGTCATCTACACGGCTGTCTT 1337

     51 eLeuValCysArgAlaCysPhe.........................P 59
        :            |||:::|||                         |
   1338 CACACCCTACTCGGCTGCCTTCCTGCTGAAGGAGACGGAAGAAGGCCCGC 1387

     59 roAspLeuGlnHisGlyTyrLeuValAlaTrpLeu...ValLeuAspTyr 74
        ||      :::   ||||||            |||   |||:::|||
   1388 CTGCTACCGAGTGTGGCTACGCCTGCCAGCCGCTGGCTGTGGTGGACCTC 1437

     75 ThrSerAspLeuLeuTyrLeuLeuAspMetValValArgPheHisThrGl 91
           |||:::::::::::::::::|||:::::::::::::|||:::|||
   1438 ATCGTGGACATCATGTTCATTGTGGACATCCTCATCAACTTCCGCACCAC 1487

     91 yPheLeuGluGlnGlyIleLeuValValAsp...LysGlyArgIleSerS 107
        :::::::::   :::      ||||||:::      |||||||||:::
   1488 CTACGTCAATGCCAACGAGGAGGTGGTCAGCCACCCCGGCCGCATCGCCG 1537

    107 erArgTyrValArgThrTrpSerPhePheLeuAspLeuAlaSerLeuMet 123
          :::|||   :::   |||   |||::::::|||:::::::::   :::
   1538 TCCACTACTTCAAGGGCTGG...TTCCTCATCGACATGGTGGCCGCCATC 1584

    124 ProThrAspValValTyrValArgLeuGlyProHis.............. 135
        |||   |||::::::               |||   :::
   1585 CCCTTCGACCTGCTCATCTTCGGCTCTGGCTCTGAGGAGCTGATCGGGCT 1634

    136 .ThrProThrLeuArgLeuAsnArgPheLeuArgAlaProArgLeuPheG 152
             |||   ||||||   |||:::::::|||:::   |||   ::::
   1635 GCTGAAGACTGCGCGGCTGCTGCGGCTGGTGCGCGTGGCGCGGAAGCTGG 1684

    152 luAlaPheAspArgThrGluThrArgThrAlaTyrProAsnAlaPheArg 168
        ::   :::::::::                          :::|||
   1685 ATCGCTACTCAGAGTAC.....................GGCGCGGCCGTG 1713

    169 IleAlaLysLeuMetLeuTyrIlePheValValIleHisTrpAsnSerCy 185
        :::      ||||||      :::   ::::::   ||||||   :::||
   1714 CTGTTCTTGCTCATGTGCACCTTTGCGCTCATCGCGCACTGGCTAGCCTG 1763

    185 sLeuTyrPheAlaLeuSerArg............TyrLeuGlyPheGlyA 198
        |:::::::::|||:::::::::            ::::::      :::|
   1764 CATCTGGTACGCCATCGGCAACATGGAGCAGCCACACATGGAC...TCAC 1810

    198 rgAspAlaTrpValTyrPro.........................Asp 205
        ||   :::|||::::::                             :::
   1811 GCATCGGCTGGCTGCACAACCTGGGCGACCAGATAGGCAAACCCTACAAC 1860

    206 ProAlaGlnProGlyPheGluArgLeuArgArgGlnTyrLeuTyrSerPh 222
          :::      |||         ::::::   :::|||:::   :::::
   1861 AGCAGCGGCCTGGGCGGCCCCTCCATCAAGGACAAGTATGTGACGGCGCT 1910

    222 eTyrPheSerThrLeuIleLeuThrThrValGly...AspThrProProP 238
        :|||||:::         ||||||:::||||||      :::       |
   1911 CTACTTCACCTTCAGCAGCCTCACCAGTGTGGGCTTCGGCAACGTCTCTC 1960

    238 roAlaArgGluGluGluTyrLeuPheMetValGlyAspPheLeuLeuAla 254
        ||      ::::::|||   :::|||   :::      :::|||::::::
   1961 CCAACACCAACTCAGAGAAGATCTTCTCCATCTGCGTCATGCTCATTGGC 2010

    255 ValMetGlyPheAlaThrIleMetGlySerMetSerSerValIleTyrAs 271
          :::   :::|||:::|||:::|||::::::|||::::::|||   ::
```

```
2011 TCCCTCATGTATGCTAGCATCTTCGGCAACGTGTCGGCCATCATCCAGCG 2060

 271 nMetAsnThrAlaAspAlaAlaPheTyrProAspHisAlaLeuValLysL 288
     ::::   ::::::   |||   ::::::   :::          |||:::: 
2061 GCTGTACTCGGGCACAGCCCGCTACCACACACAGATGCTGCGGGTGCGGG 2110

 288 ysTyrMetLysLeuGlnHisValAsnArgLysLeuGluArgArgValIle 304
     ::::::::::::::::::::::::   :::   |||:::::::|||::: 
2111 AGTTCATCCGCTTCCACCAGATCCCCAATCCCCTGCGCCAGCGCCTCGAG 2160

 305 AspTrpTyrGlnHisLeuGlnIleAsnLysLysMetThrAsnGluVal.. 320
     :::::::::||||||               :::   ||||||   :::    
2161 GAGTACTTCCAGCACGCCTGG.........TCCTACACCAACGGCATCGA 2201

 321 .......AlaIleLeuGlnHisLeuProGluArgLeuArgAlaGluValA 335
           |||:::|||:::   :::||||||   |||:::|||:::::::  
2202 CATGAACGCGGTGCTGAAGGGCTTCCCTGAGTGCCTGCAGGCTGACATCT 2251

 335 laValSerValHisLeuSerThrLeuSerArgValGlnIlePheGlnAsn 351
     :::::   ::::::   |||   |||::::::   :::   |||::::::  
2252 GCCTGCACCTGAACCGCTCACTGCTGCAGCACTGCAAACCCTTCCGAGGG 2301

 352 CysGluAlaSerLeuLeuGluGluLeuValLeuLysLeuGlnProGlnTh 368
     :::      :::   |||:::   |||::::::|||::::::  
2302 GCCACCAAGGGCTGCCTTCGGGCCCTGGCCATGAAGTTCAAGACCACACA 2351

 368 rTyrSerProGlyGluTyrValCysArgLysGlyAspIleGlyGlnGluM 385
           ||||||:::   :::   :::   ||||||:::         :  
2352 TGCACCGCCAGGGGACACACTGGTGCATGCTGGGGACCTGCTCACCGCCC 2401

 385 etTyrIleIleArgGluGlyGlnLeuAlaValValAlaAspAspGlyIle 401
     ::|||:::|||   :::|||::::::   ::::::      |||   :::  
2402 TGTACTTCATCTCCCGGGGCTCCATCGAGATCCTGCGGGGCGACGTCGTC 2451

 402 ThrGlnTyrAlaValLeuGlyAlaGlyLeuTyrPheGlyGlu...IleSe 417
     :::      |||:::||||||   :::      |||||||||   :::::  
2452 GTG......GCCATCCTGGGGAAGAATGACATCTTTGGGGAGCCTCTGAA 2495

 417 rIleIleAsnIleLysGlyAsnMetSerGlyAsnArgArgThrAlaAsnI 434
     ::::            |||:::   :::|||               ::::  
2496 CCTGTATGCAAGGCCTGGCAAGTCGAACGGG.............GATG 2530

 434 leLysSerLeuGlyTyrSerAspLeuPheCysLeuSerLysGluAspLeu 450
     ::::::::|||   |||   ||||||      :::   ::::::||||||  
2531 TGCGGGCCCTCACCTACTGTGACCTACACAAGATCCATCGGGACGACCTG 2580

 451 ArgGluValLeuSerGluTyrProGlnAlaGlnThrIleMetGluGluLy 467
        ||||||||||:::   ||||||:::         :::   ::::::  
2581 CTGGAGGTGCTGGACATGTACCCTGAG.........TTCTCCGACCACTT 2621

 467 sGlyArgGluIleLeuLeuLysMetAsnLysLeuAspValAsnAlaGluA 484
           ::::::   :::   :::|||  .    |||:::|||  
2622 CTGGTCCAGCCTGGAGATCACCTTCAACCTGCGAGATACCAACATGATCC 2671

 484 laAlaGluIleAlaLeuGlnGluAlaThrGluSerArgLeuArgGlyLeu 500
       ::::::   :::      |||
2672 CGGGCTCCCCCGGCAGTACGGAG.......................... 2694

 501 AspGlnGlnLeuAspAspLeuGlnThrLysPheAlaArgLeuLeuAlaGl 517
                    |||:::      |||:::|||
2695 .................TTAGAGGGTGGCTTCAGTCGG........... 2715

 517 uLeuGluSerSerAlaLeuLysIleAlaTyrArgIleGluArgLeuGluT 534
                                  |||   ::::::|||:::: 
2716 ..........................CAACGCAAGCGCAAGTTGTCCT 2737

 534 rpGlnThrArgGluTrpProMetProGluAspLeuAlaGluAlaAspAsp 550
     :::::   |||                |||:::   :::|||::::::  
2738 TCCGCAGGCGCACGGACAAGGACACGGAGCAGCCAGGGGAGGTGTCGGCC 2787

 551 GluGlyGluProGluGluGlyThrSerLysAspGluGluGlyArg..... 565
        |||      :::   |||::::::   :::::::::||||||  
2788 TTGGGGCCGGGCCGGGCGGGGGCAGGGCCGAGTAGCCGGGGCCGGCCGGG 2837

      566 ...............AlaSerGlnGluGlyProProGlyProGlu 575
                              :::   ::::::||||||   :::|||||| 
     2838 GGGGCCGTGGGGGGAGAGCCCGTCCAGTGGCCCCTCCAGCCCTGAG 2883
```

```
seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-956-242-3

seq_documentation_block:
; Sequence 3, Application US/08956242C
; Patent No. 5986081
; GENERAL INFORMATION:
;  APPLICANT: Ganetzky, Barry S.
;  APPLICANT: Titus, Steven A.
;  TITLE OF INVENTION: Polynucleotides Encoding Herg-3
;  FILE REFERENCE: 960296.94550
;  CURRENT APPLICATION NUMBER: US/08/956,242C
;  CURRENT FILING DATE: 1997-10-22
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 3142
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (34)..(2700)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (3)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (17)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (31)..(32)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (432)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (575)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1346)..(1348)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1733)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1735)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1737)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1744)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1746)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1758)
```

```
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1773)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1798)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1812)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1830)..(1831)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1855)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1862)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1864)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1895)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1900)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1903)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1915)..(1916)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1920)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1934)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1943)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1971)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1990)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (2001)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (2013)..(2019)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (3124)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (3135)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (3141)
;   OTHER INFORMATION: Unidentified at time of filing
US-08-956-242-3


alignment_scores:
            Quality:  370.50              Length:    506
              Ratio:  1.291                 Gaps:     16
 Percent Similarity:  56.719   Percent Identity: 22.727

alignment_block:
 US-09-927-267-1 x US-08-956-242-3   ..

 Align seg 1/1  to: US-08-956-242-3  from: 1  to: 3142

      36 Trp...TrpLeuAsnThrMetValPheProValMetTyrAsnLeuIleIl 51
         |||   ||||||      :::::::::   |||:::|||:::   |||::
     190 TGGGACTGGCTT......ATCCTGCTGTTGGTCATATACACTGCTATATT 233

      51 eLeuValCysArgAlaCysPhe.......................ProA 60
         :           |||:::|||
     234 TACTCCCTACTCTGCAGCCTTCCTCCTCAATGACAGAGAAGAACAGAAGA 283

      60 spLeuGlnHisGlyTyrLeuValAlaTrpLeu...ValLeuAspTyrThr 75
           :::   ||||||      :::   |||   |||:::|||
     284 GACGAGAATGTGGCTATTCTTGTAGCCCTTTGAATGTGGTAGACTTGATT 333

      76 SerAspLeuLeuTyrLeuLeuAspMetValValArgPheHisThrGlyPh 92
         |||:::::::::::::::|||:::::::::::::|||:::|||   ::
     334 GTGGATATTATGTTTATCATAGATATTTTAATAAACTTCAGAACAACATA 383

      92 eLeuGluGlnGlyIleLeuValVal...AspLysGlyArgIleSerSerA 108
         ::::::::|||:::      ||||||   |||   ::::::|||:::::::
     384 TGTTAATCAGAATGAAGAAGTGGTAAGTGATCCCGCCAAAATAGCAATNC 433

     108 rgTyrValArgThrTrpSerPhePheLeuAspLeuAlaSerLeuMetPro 124
         ::|||   :::   |||   |||:::::|||:::::::::   :::|||
     434 ACTACTTCAAAGGCTGG...TTCCTGATTGACATGGTTGCAGCAATTCCT 480

     125 ThrAspValValTyrValArgLeuGlyProHisThrProThr........ 138
         |||::::::            |||            |||
     481 TTTGACTTGCTGATTTTTGGATCAGGTTCTGATGAGACAACAACATTAAT 530

     139 .......LeuArgLeuAsnArgPheLeuArgAlaProArgLeuPheGluA 153
               |||:::      |||:::||||||     |||:::   :::
     531 TGGTCTTTTGAAGACTGCCCGACTCCTCCGTCTTGTGCGCGTGGNCAGGA 580

     153 laPheAspArgThrGluThrArgThrAlaTyrProAsnAlaPheArgIle 169
           :::||||||   :::         |||                 :::
     581 AACTGGATCGATATTCAGAATATGGCGCT.............GCTGTT 615

     170 AlaLysLeuMetLeuTyrIlePheValVal...IleHisTrpAsnSerCy 185
           |||   :::   ||||||::::::      ||||||   :::||
     616 CTAATGCTCTCAATGTGCATCTTTGCCCTGAATGCACACTGGCTGGCTTG 665

     185 sLeuTyrPheAlaLeuSerArg...........TyrLeuGlyPheGlyA 198
         |::::::::::|||:::::::::          ||||||        :
     666 CATTTGGTATGCGATTGGGAATGTAGAAAGGCCTTACCTG...ACTGACA 712

     198 rgAspAlaTrpVal.........................TyrPro 204
         ::   :::|||:::                          |||
     713 AAATCGGATGGTTGGATTCCTTAGGACAGCAAATTGGGAAACGTTACAAT 762
```

```
  205 AspProAlaGlnProGlyPheGluArgLeuArgArgGlnTyrLeuTyrSe 221
      |||       :::   :::         ::::::   :::|||:::   ::
  763 GACAGTGACTCAAGTTCTGGACCATCCATTAAAGACAAATACGTCACAGC 812

  221 rPheTyrPheSerThrLeuIleLeuThrThrValGly...AspThrProP 237
      ::::||||||:::          ||||||:::||||||      :::
  813 ACTTTATTTTACCTTCAGCAGTTTAACCAGTGTAGGATTCGGGAATGTGT 862

  237 roProAlaArgGluGluGluTyrLeuPheMetValGlyAspPheLeuLeu 253
        |||       :::::|||   :::|||   :::       :::|||:::
  863 CTCCTAACACGAATTCGGAGAAAATCTTTTCAATTTGTGTCATGTTGATT 912

  254 AlaValMetGlyPheAlaThrIleMetGlySerMetSerSerValIleTy 270
      :::   :::   :::|||:::|||:::|||:::::::|||:::::::|||
  913 GGCTCACTAATGTATGCAAGCATTTTTGGGAATGTATCTGCAATTATCCA 962

  270 rAsnMetAsnThrAlaAspAlaAlaPheTyrProAspHisAlaLeuValL 287
       ::::::   ::::::   |||   ::::::   :::          ||||
  963 AAGACTATACTCGGGAACTGCCAGGTACCACATGCAGATGCTGCGAGTAA 1012

  287 ysLysTyrMetLysLeuGlnHisValAsnArgLysLeuGluArgArgVal 303
      ||:::::::::::::::::::::::::   :::   |||::::::|||:::
 1013 AAGAGTTCATTCGCTTTCACCAAATCCCCAACCCTCTGAGGCAACGTCTT 1062

  304 IleAspTrpTyrGlnHisLeuGlnIleAsnLysLysMetThrAsnGluVa 320
        :::::::::||||||           :::       :::
 1063 GAAGAATATTTCCAGCACGCATGGACTTACACCAATGGCATTGACATGAA 1112

  320 lAlaIleLeuGlnHisLeuProGluArgLeuArgAlaGluValAlaValS 337
        :::|||:::   :::||||||   |||:::|||:::::::::::::
 1113 CATGGTCCTAAAGGGTTTCCCAGAATGCTTACAAGCAGACATTTGTCTAC 1162

  337 erValHisLeuSerThrLeuSerArgValGlnIlePheGlnAsnCysGlu 353
      ::::::   :::   |||:::::   :::   |||::::::::::::
 1163 ATCTCAACCAGACATTGCTGCAAAACTGCAAAGCCTTTCGGGGGGCAAGT 1212

  354 AlaSerLeuLeuGluGluLeuValLeuLysLeuGlnProGlnThrTyrSe 370
        :::>  |||:::   |||:::::::|||::::::
 1213 AAAGGTTGCCTTAGAGCTTTGGCAATGAAGTTCAAAACCACCCATGCACT 1262

  370 rProGlyGluTyrValCysArgLysGlyAspIleGlyGlnGluMetTyrI 387
        |||:::   :::   :::   ||||||:::         :::|||:
 1263 CCAAGGAGACACCCTCGTTCACTGTGGGGATGTCCTCACTGCACTTTATT 1312

  387 leIleArgGluGlyGlnLeuAlaValValAlaAspAspGlyIleThrGln 403
      :::::   :::|||::::::   :::                ::::::
 1313 TCTTATCCAGAGGCTCCATTGAAATCTCAAAGANNNGACTTGTGGTG... 1359

  404 TyrAlaValLeuGlyAlaGlyLeuTyrPheGlyGluIleSerIleIleAs 420
        |||:::||||||   :::      |||||||||:::      :::
 1360 ...GCTATTCTGGGAAAAAATGATATATTTGGAGAAATGGTTCATCTTTA 1406

  420 nIleLysGlyAsnMetSerGlyAsnArgArgThrAlaAsnIleLysSerL 437
        |||   :::   |||            :::|||:::::::::::::|
 1407 TGCCAAACCTGGAAAGTCT.....AATGCAGATGTAAGAGCCC 1444

  437 euGlyTyrSerAspLeuPheCysLeuSerLysGluAspLeuArgGluVal 453
      ||   |||   ||||||      :::::::::||||||||||   ||||||
 1445 TCACATACTGTGACTTGCATAAGATTCAGCGAGAAGACTTGTTAGAGGTT 1494

  454 LeuSerGluTyrProGln..................................... 459
      |||:::   ||||||:::
 1495 TTGGATATGTATCCTGAGTTTTCTGATCACTTTCTAACAAACCTAGAGTT 1544

  460 .....................................AlaG 461
                                          :::|
 1545 GACTTTCAACCTAAGGCATGAGAGCGCAAAGGCTGATCTCCTACGATCAC 1594

  461 lnThrIleMetGluGluLysGlyArgGluIleLeuLeuLysMetAsnLys 477
      ||:::::::   :::::::::|||   :::       |||:::   :::|||
 1595 AATCCATGAATGATTCAGAAGGAGACAACTGTAAACTAAGAAGAAGGAAA 1644

  478 LeuAspValAsnAlaGlu 483
```

```
          |||:::   ::::::|||
     1645 TTGTCATTTGAAAGTGAA 1662

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-351-215-3

seq_documentation_block:
; Sequence 3, Application US/09351215
; Patent No. 6087488
; GENERAL INFORMATION:
;  APPLICANT: Ganetzky, Barry S.
;  APPLICANT: Titus, Steven A.
;  TITLE OF INVENTION: Polynucleotides Encoding Herg-3
;  FILE REFERENCE: 960296.94550
;  CURRENT APPLICATION NUMBER: US/09/351,215
;  CURRENT FILING DATE: 1999-07-12
;  EARLIER APPLICATION NUMBER: 08/956,242
;  EARLIER FILING DATE: 1997-10-22
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 3142
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (34)..(2700)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (3)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (17)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (31)..(32)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (432)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (575)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1346)..(1348)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1733)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1735)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1737)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1744)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1746)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1758)
```

```
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1773)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1798)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1812)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1830)..(1831)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1855)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1862)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1864)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1895)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1900)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1903)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1915)..(1916)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1920)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1934)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1943)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1971)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1990)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (2001)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (2013)..(2019)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (3124)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (3135)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (3141)
;   OTHER INFORMATION: Unidentified at time of filing
US-09-351-215-3


alignment_scores:
            Quality:  370.50              Length:    506
              Ratio:  1.291                 Gaps:     16
 Percent Similarity:  56.719   Percent Identity: 22.727

alignment_block:
 US-09-927-267-1 x US-09-351-215-3   ..

 Align seg 1/1  to: US-09-351-215-3  from: 1  to: 3142

      36 Trp...TrpLeuAsnThrMetValPheProValMetTyrAsnLeuIleIl 51
         |||   ||||||      :::::::::   |||:::|||:::   |||::
     190 TGGGACTGGCTT......ATCCTGCTGTTGGTCATATACACTGCTATATT 233

      51 eLeuValCysArgAlaCysPhe.....................ProA 60
         :          |||:::|||
     234 TACTCCCTACTCTGCAGCCTTCCTCCTCAATGACAGAGAAGAACAGAAGA 283

      60 spLeuGlnHisGlyTyrLeuValAlaTrpLeu...ValLeuAspTyrThr 75
           :::   ||||||      :::   |||   |||:::|||
     284 GACGAGAATGTGGCTATTCTTGTAGCCCTTTGAATGTGGTAGACTTGATT 333

      76 SerAspLeuLeuTyrLeuLeuAspMetValValArgPheHisThrGlyPh 92
         |||:::::::::::::::|||:::::::::::::|||:::|||   ::
     334 GTGGATATTATGTTTATCATAGATATTTTAATAAACTTCAGAACAACATA 383

      92 eLeuGluGlnGlyIleLeuValVal...AspLysGlyArgIleSerSerA 108
         :::::::|||:::      ||||||   |||   :::::::|||:::::::
     384 TGTTAATCAGAATGAAGAAGTGGTAAGTGATCCCGCCAAAATAGCAATNC 433

     108 rgTyrValArgThrTrpSerPhePheLeuAspLeuAlaSerLeuMetPro 124
         ::|||   :::   |||   |||:::::|||:::::::::   :::|||
     434 ACTACTTCAAAGGCTGG...TTCCTGATTGACATGGTTGCAGCAATTCCT 480

     125 ThrAspValValTyrValArgLeuGlyProHisThrProThr........ 138
         |||::::::           |||           |||
     481 TTTGACTTGCTGATTTTTGGATCAGGTTCTGATGAGACAACAACATTAAT 530

     139 .......LeuArgLeuAsnArgPheLeuArgAlaProArgLeuPheGluA 153
               |||:::      |||:::|||||      |||:::   :::
     531 TGGTCTTTTGAAGACTGCCCGACTCCTCCGTCTTGTGCGCGTGGNCAGGA 580

     153 laPheAspArgThrGluThrArgThrAlaTyrProAsnAlaPheArgIle 169
           :::||||||   :::         |||                 :::
     581 AACTGGATCGATATTCAGAATATGGCGCT.............GCTGTT 615

     170 AlaLysLeuMetLeuTyrIlePheValVal...IleHisTrpAsnSerCy 185
             |||   :::   ||||||:::::      ||||||   :::||
     616 CTAATGCTCTCAATGTGCATCTTTGCCCTGAATGCACACTGGCTGGCTTG 665

     185 sLeuTyrPheAlaLeuSerArg...........TyrLeuGlyPheGlyA 198
         |:::::::::|||:::::::::          ||||||        :
     666 CATTTGGTATGCGATTGGGAATGTAGAAAGGCCTTACCTG...ACTGACA 712

     198 rgAspAlaTrpVal.........................TyrPro 204
         ::   :::|||:::                          |||
     713 AAATCGGATGGTTGGATTCCTTAGGACAGCAAATTGGGAAACGTTACAAT 762
```

```
 205 AspProAlaGlnProGlyPheGluArgLeuArgArgGlnTyrLeuTyrSe 221
 763 GACAGTGACTCAAGTTCTGGACCATCCATTAAAGACAAATACGTCACAGC 812

 221 rPheTyrPheSerThrLeuIleLeuThrThrValGly...AspThrProP 237
 813 ACTTTATTTTACCTTCAGCAGTTTAACCAGTGTAGGATTCGGGAATGTGT 862

 237 roProAlaArgGluGluGluTyrLeuPheMetValGlyAspPheLeuLeu 253
 863 CTCCTAACACGAATTCGGAGAAAATCTTTTCAATTTGTGTCATGTTGATT 912

 254 AlaValMetGlyPheAlaThrIleMetGlySerMetSerSerValIleTy 270
 913 GGCTCACTAATGTATGCAAGCATTTTTGGGAATGTATCTGCAATTATCCA 962

 270 rAsnMetAsnThrAlaAspAlaAlaPheTyrProAspHisAlaLeuValL 287
 963 AAGACTATACTCGGGAACTGCCAGGTACCACATGCAGATGCTGCGAGTAA 1012

 287 ysLysTyrMetLysLeuGlnHisValAsnArgLysLeuGluArgArgVal 303
1013 AAGAGTTCATTCGCTTTCACCAAATCCCCAACCCTCTGAGGCAACGTCTT 1062

 304 IleAspTrpTyrGlnHisLeuGlnIleAsnLysLysMetThrAsnGluVa 320
1063 GAAGAATATTTCCAGCACGCATGGACTTACACCAATGGCATTGACATGAA 1112

 320 lAlaIleLeuGlnHisLeuProGluArgLeuArgAlaGluValAlaValS 337
1113 CATGGTCCTAAAGGGTTTCCCAGAATGCTTACAAGCAGACATTTGTCTAC 1162

 337 erValHisLeuSerThrLeuSerArgValGlnIlePheGlnAsnCysGlu 353
1163 ATCTCAACCAGACATTGCTGCAAAACTGCAAAGCCTTTCGGGGGGCAAGT 1212

 354 AlaSerLeuLeuGluGluLeuValLeuLysLeuGlnProGlnThrTyrSe 370
1213 AAAGGTTGCCTTAGAGCTTTGGCAATGAAGTTCAAAACCACCCATGCACT 1262

 370 rProGlyGluTyrValCysArgLysGlyAspIleGlyGlnGluMetTyrI 387
1263 CCAAGGAGACACCCTCGTTCACTGTGGGGATGTCCTCACTGCACTTTATT 1312

 387 leIleArgGluGlyGlnLeuAlaValValAlaAspAspGlyIleThrGln 403
1313 TCTTATCCAGAGGCTCCATTGAAATCTCAAAGANNNGACTTGTGGTG... 1359

 404 TyrAlaValLeuGlyAlaGlyLeuTyrPheGlyGluIleSerIleIleAs 420
1360 ...GCTATTCTGGGAAAAAATGATATATTTGGAGAAATGGTTCATCTTTA 1406

 420 nIleLysGlyAsnMetSerGlyAsnArgArgThrAlaAsnIleLysSerL 437
1407 TGCCAAACCTGGAAAGTCT.....AATGCAGATGTAAGAGCCC 1444

 437 euGlyTyrSerAspLeuPheCysLeuSerLysGluAspLeuArgGluVal 453
1445 TCACATACTGTGACTTGCATAAGATTCAGCGAGAAGACTTGTTAGAGGTT 1494

 454 LeuSerGluTyrProGln..... 459
1495 TTGGATATGTATCCTGAGTTTTCTGATCACTTTCTAACAAACCTAGAGTT 1544

 460 .....AlaG 461
1545 GACTTTCAACCTAAGGCATGAGAGCGCAAAGGCTGATCTCCTACGATCAC 1594

 461 lnThrIleMetGluGluLysGlyArgGluIleLeuLeuLysMetAsnLys 477
1595 AATCCATGAATGATTCAGAAGGAGACAACTGTAAACTAAGAAGAAGGAAA 1644

 478 LeuAspValAsnAlaGlu 483
```

```
1645 TTGTCATTTGAAAGTGAA 1662

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-600-776-10

seq_documentation_block:
; Sequence 10, Application US/09600776
; Patent No. 6326168
; GENERAL INFORMATION:
;  APPLICANT:  Yamanouchi Pharmaceutical Co., Ltd.
;  TITLE OF INVENTION:  A novel potassium channel protein
;  FILE REFERENCE:  Y9903-PCT
;  CURRENT APPLICATION NUMBER: US/09/600,776
;  CURRENT FILING DATE:  2000-07-21
;  PRIOR APPLICATION NUMBER:  JP P1998-011434
;  PRIOR FILING DATE:  1998-01-23
;  PRIOR APPLICATION NUMBER:  JP P1998-346198
;  PRIOR FILING DATE:  1998-12-04
;  NUMBER OF SEQ ID NOS:  12
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 10
;   LENGTH: 3736
;   TYPE: DNA
;   ORGANISM: Rattus sp.
US-09-600-776-10


alignment_scores:
            Quality:  362.00              Length:    448
              Ratio:  1.356                 Gaps:     12
 Percent Similarity:  59.598   Percent Identity: 25.000

alignment_block:
 US-09-927-267-1 x US-09-600-776-10   ..

 Align seg 1/1  to: US-09-600-776-10  from: 1  to: 3736

  37 TrpLeuAsnThrMetValPheProValMetTyrAsnLeuIleIleLeuVa 53
 823 TGGGACGGTCTCATCCTTCTCGCTACGTTCTACGTCGCGGTCACCGTCCC 872

  53 lCysArgAlaCysPheProAspLeuGlnHisGlyTyrLeuValAla.... 68
 873 TTACAACGTCTGCTTCGCTGGTGATGACGACACCCCCATCACGTCCCGAC 922

  69 ..TrpLeuValLeuAspTyrThrSerAspLeuLeuTyrLeuLeuAspMet 84
 923 ACACCCTTGTCAGTGACATCGCTGTGGAGATGCTCTTCATCCTGGACATC 972

  85 ValValArgPheHisThrGlyPheLeuGluGlnGlyIleLeuValValAs 101
 973 ATCTTGAACTTCCGCACCACCTACGTGTCCCAGTCGGGCCAGGTGGTTTC 1022

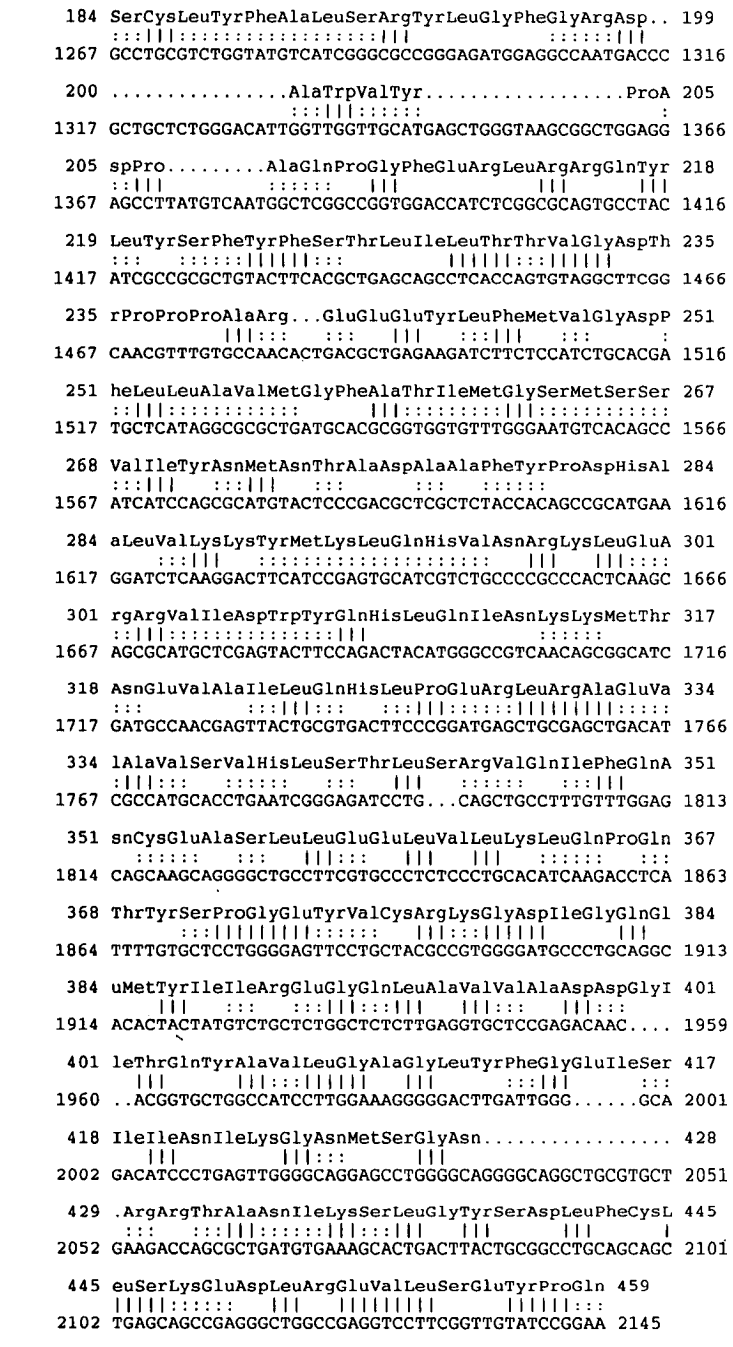
 101 pLysGlyArg...IleSerSerArgTyrValArgThrTrpSerPhePheL 117
1023 TGCTCCTCGGTCCATTGGCCTCCACTACCTGGCCACCTGG...TTCTTCG 1069

 117 euAspLeuAlaSerLeuMetProThrAspValValTyrValArgLeuGly 133
1070 TGGACCTCATTGCTGCTTTGCCCTTTGACCTGCTGTATGTCTTCAACATC 1119

 134 ProHisThrProThrLeuArgLeuAsnArgPheLeuArgAlaProArgLe 150
1120 ACTGTGACCTCGCTGGTACATCTGCTGAAAACCGTGCGGCTCCTGCGGTT 1169

 150 uPheGluAlaPheAspArgThrGluThrArgThrAlaTyrProAsnAlaP 167
1170 GCTGAGGCTGCTGCAGAAGCTAGAG...CGGTACTCTCAGTGCAGCGCGG 1216

 167 heArgIleAlaLysLeuMetLeuTyrIlePheValValIleHisTrpAsn 183
1217 TGGTGCTCACGCTGCTCATGTCCGTCTTTGCACTCCTTGCCCACTGGATG 1266
```



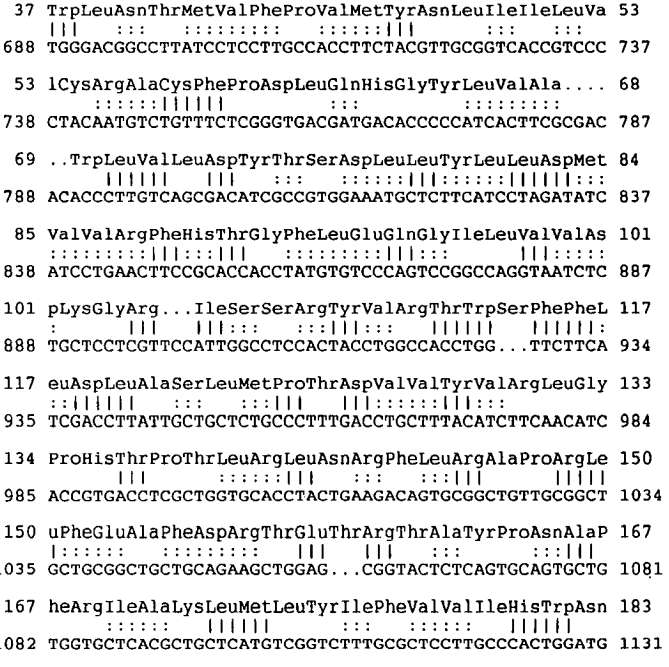
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-600-776-5

```
seq_documentation_block:
; Sequence 5, Application US/09600776
; Patent No. 6326168
; GENERAL INFORMATION:
;  APPLICANT:  Yamanouchi Pharmaceutical Co., Ltd.
;  TITLE OF INVENTION:  A novel potassium channel protein
;  FILE REFERENCE:  Y9903-PCT
;  CURRENT APPLICATION NUMBER: US/09/600,776
;  CURRENT FILING DATE:  2000-07-21
;  PRIOR APPLICATION NUMBER:  JP P1998-011434
;  PRIOR FILING DATE:  1998-01-23
;  PRIOR APPLICATION NUMBER:  JP P1998-346198
;  PRIOR FILING DATE:  1998-12-04
;  NUMBER OF SEQ ID NOS:  12
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 3064
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (4)..(3057)
US-09-600-776-5


alignment_scores:
            Quality:  361.00              Length:    451
              Ratio:  1.352                 Gaps:     12
 Percent Similarity:  59.202   Percent Identity: 24.612

alignment_block:
 US-09-927-267-1 x US-09-600-776-5   ..

 Align seg 1/1  to: US-09-600-776-5  from: 1  to: 3064
```



```
1132 GCCTGCATCTGGTATGTCATCGGGCGCCGGGAGATGGAGGCCAATGACCC 1181

 200 .................AlaTrpVal........................ 202
                     :::|||:::
1182 GCTGCTCTGGGACATTGGCTGGTTGCATGAGTTGGGCAAGCGTCTGGAGG 1231

 203 .....TyrProAspProAlaGlnProGlyPheGluArgLeuArgArgGln 217
          |||   :::   :::      |||   :::|||   |||
1232 TGCCCTATGTCAATGGCTCGGTGGGCGGCCCATCACGG...CGCAGCGCC 1278

 218 TyrLeuTyrSerPheTyrPheSerThrLeuIleLeuThrThrValGlyAs 234
     |||:::   ::::::||||||:::         ||||||:::||||||
1279 TACATCGCGGCACTGTACTTCACTCTAAGCAGCCTCACCAGTGTGGGCTT 1328

 234 pThrProProProAlaArg...GluGluGluTyrLeuPheMetValGlyA 250
               |||:::   :::   |||   :::|||   :::
1329 TGGCAACGTGTGTGCCAACACCGACGCGGAGAAGATCTTCTCCATCTGCA 1378

 250 spPheLeuLeuAlaValMetGlyPheAlaThrIleMetGlySerMetSer 266
       :::|||:::::::::::::      |||:::::::::|||:::::::::
1379 CGATGCTCATAGGCGCCCTGATGCACGCTGTGGTGTTCGGGAACGTGACA 1428

 267 SerValIleTyrAsnMetAsnThrAlaAspAlaAlaPheTyrProAspHi 283
     :::::::|||   :::|||   :::      :::   ::::::
1429 GCCATCATCCAGCGCATGTACTCGCGCCGCTCGCTCTACCACAGCCGCAT 1478

 283 sAlaLeuValLysLysTyrMetLysLeuGlnHisValAsnArgLysLeuG 300
           :::|||   :::::::::::::::::::::   |||   |||:
1479 GAAGGACCTCAAGGACTTCATCCGTGTGCACCGCCTGCCGCGGCCGCTCA 1528

 300 luArgArgValIleAspTrpTyrGlnHisLeuGlnIleAsnLysLysMet 316
     :::::|||:::::::::::::::|||           ::::::
1529 AGCAGCGCATGCTCGAATACTTCCAGACCACGTGGGCCGTCAACAGCGGC 1578

 317 ThrAsnGluValAlaIleLeuGlnHisLeuProGluArgLeuArgAlaGl 333
        :::         :::|||:::   :::|||::::::|||||||::
1579 ATCGACGCCAACGAGTTACTGCGTGACTTCCCAGACGAGCTGAGAGCTGA 1628

 333 uValAlaValSerValHisLeuSerThrLeuSerArgValGlnIlePheG 350
     ::::|||:::   ::::::   :::   |||   ::::::   :::|||
1629 CATTGCTATGCACCTGAATCGGGAGATCCTG...CAGCTGCCGTTGTTCG 1675

 350 lnAsnCysGluAlaSerLeuLeuGluGluLeuValLeuLysLeuGlnPro 366
           ::::::   :::   |||:::   |||   |||   ::::::
1676 GGGCAGCGAGCAGGGGCTGCCTGCGGGCCCTATCGCTGCACATCAAGACC 1725

 367 GlnThrTyrSerProGlyGluTyrValCysArgLysGlyAspIleGlyGl 383
     :::      :::||||||||||||:::   |||:::||||||      ||
1726 TCGTTCTGCGCTCCGGGCGAGTACCTGTTGCGCCGTGGGGATGCCCTGCA 1775

 383 nGluMetTyrIleIleArgGluGlyGlnLeuAlaValValAlaAspAspG 400
     |      |||   :::   :::|||:::|||   |||:::   |||:::
1776 GGCACATTACTATGTCTGCTCCGGCTCGCTTGAGGTGCTCCGAGACAACA 1825

 400 lyIleThrGlnTyrAlaValLeuGlyAlaGlyLeuTyrPheGlyGluIle 416
       :::         |||:::||||||   |||      :::|||
1826 TGGTG......CTGGCCATCCTGGGGAAGGGGGACCTGATTGGA...... 1863

 417 SerIleIleAsnIleLysGlyAsnMetSerGly.................. 427
     :::   |||         |||:::      |||
1864 GCAGATATCCCTGAGCCGGGGCAGGAGCCTGGGTTGGGAGCAGACCCAAA 1913

 428 .......AsnArgArgThrAlaAsnIleLysSerLeuGlyTyrSerAspL 442
             :::   :::|||::::::|||:::|||   |||      |
1914 CTTCGTGCTAAAGACCAGTGCTGATGTGAAAGCTCTGACCTACTGTGGCC 1963

 442 euPheCysLeuSerLysGluAspLeuArgGluValLeuSerGluTyrPro 458
     ||      ||||||::::::   |||   |||||||||      ||||||
1964 TGCAGCAGCTGAGCAGCCGAGGGCTGGCTGAGGTCCTGAGGCTCTATCCT 2013

 459 Gln 459
     :::
```

2014 GAG 2016

```
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-172-422-3

seq_documentation_block:
; Sequence 3, Application US/09172422A
; Patent No. 6300485
; GENERAL INFORMATION:
;  APPLICANT: Adams, Arwen E.
;  APPLICANT: Chiu, Choi Ying
;  APPLICANT: Duhl, David
;  APPLICANT: Gorman, Susan W.
;  APPLICANT: Leng, Song
;  APPLICANT: Sheffield, Val
;  APPLICANT: Welch, Juliet
;  TITLE OF INVENTION: MYOSIN IXA AND CYCLIC NUCLEOTIDE GATED
;  TITLE OF INVENTION: CHANNEL-15 (CNGC-15) POLYNUCLEOTIDES, POLYPEPTIDES,
;  TITLE OF INVENTION: COMPOSITIONS, METHODS, AND USES THEREOF
;  FILE REFERENCE: 200130.442
;  CURRENT APPLICATION NUMBER: US/09/172,422A
;  CURRENT FILING DATE: 1998-10-14
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
;   LENGTH: 1307
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1307)
;   OTHER INFORMATION: n = A,T,C or G
US-09-172-422-3


alignment_scores:
            Quality:  352.50              Length:    403
              Ratio:  1.469                 Gaps:     11
 Percent Similarity:  59.553   Percent Identity: 23.573

alignment_block:
 US-09-927-267-1 x US-09-172-422-3   ..

 Align seg 1/1  to: US-09-172-422-3  from: 1  to: 1307

      26 ValLeuAspProSerGlyAspTyrTyrTyrTrpTrpLeuAsnThrMetVa 42
         ::::::   |||   :::|||:::   :::::::|||      ||||||::
     102 ATTATCCACCCCTACAGTGACTTCAGATTTTACTGGGACCTGACCATGCT 151

      42 lPheProValMetTyrAsnLeuIleIleLeuValCysArgAlaCysPheP 59
         ::::   ::::::   ||||||||||||:::              |||
     152 GCTGCTGATGGTGGGAAACCTGATTATCATTCCTGTGGGCATCACCTTC. 200

      59 roAspLeuGlnHisGlyTyrLeuValAlaTrpLeuValLeuAspTyrThr 75
              ::::::         :::   |||:::|||::::::   :::
     201 .....TTCAAGGATGAGAACACCACACCCTGGATTGTCTTCAATGTGGTG 245

      76 SerAspLeuLeuTyrLeuLeuAspMetValValArgPheHisThrGlyPh 92
         ||||||   ::::::|||:::|||:::|||   :::|||:::||||||::
     246 TCAGACACATTCTTCCTCATCGACTTGGTCTTCAACTTCCGCACAGGGAT 295

      92 eLeu......GluGlnGlyIleLeuValValAspLysGlyArgIleSerS 107
         ::::      ::::::      ::::::::::|||      ||||||:::
     296 CGTGGTGGAGGACAACACAGAGATCATCCTGGACCCGCAGCGGATTAAAA 345

     107 erArgTyrValArgThrTrpSerPhePheLeuAspLeuAlaSerLeuMet 123
           :::|||::::::::::|||   |||:::::::|||:::   |||   :::
     346 TGAAGTACCTGAAAAGCTGG...TTCATGGTAGATTTCATTTCCTCCATC 392

     124 .ProThrAspValValTyr.............................. 129
          |||:::::::   ::::::
     393 CCCCGTGGAAAACATCTTCCTCATTGTGGAGACAGGCATCGACTCGGAGG 442

     130 ...................ValArgLeuGlyProHisThrProThrLeu 139
```

```
                                :::|||:::         |||      |||
    443 TCTACAAGACTGCCCGGGCCCTGCGCATTGTCCGCTTCACGAAGATCCTC 492

    140 ArgLeuAsnArgPheLeuArgAlaProArgLeuPheGluAlaPheAspAr 156
           |||   |||:::||||||      ||||||:::::   :::   ::
    493 AGCCTCTTACGCCTGTTACGCCTCTCCCGCCTCATTCGATATATTCACCA 542

    156 gThrGlu..............ThrArgThrAlaTyrProAsnAlaPheA 168
        :   |||               |||                  :::   |
    543 GTGGGAAGAGATCTTCCACATGACCTACGACCTGGCCAGCGCCGTGGTGC 592

    168 rgIleAlaLysLeuMetLeuTyrIlePheValValIleHisTrpAsnSer 184
        |||||::::::|||:::      ::::::::::::   ||||||::::::
    593 GCATCGTGAACCTCATCGGCATGATGCTCCTGCTCTGCCACTGGGACGGC 642

    185 CysLeuTyrPheAlaLeuSerArgTyrLeuGlyPheGlyArgAspAlaTr 201
        ||||||   |||   :::               |||      |||:::||
    643 TGCCTGCAGTTCCTGGTACCCATGCTACAGGACTTCCCTGACGACTGCTG 692

    201 pValTyrProAspProAlaGlnProGlyPheGluArgLeu..... 214
        ||||                 ::::::::::::::
    693 GGTG.....................TCCATCAACAACATGGTGAACAACT 721

    215 .....ArgArgGlnTyrLeuTyrSerPheTyr......PheSerThrLeu 227
              :::|||||||   |||:::::::::::      :::|||   :::
    722 CCTGGGGGAAGCAGTACTCCTACGCGCTCTTCAAGGCCATGAGCCACATG 771

    228 IleLeuThrThrValGlyAspThrProProProAlaArgGluGluGluTy 244
        :::           |||          |||   :::   ::::::   ::
    772 CTGTGCATCGGCTACGGGCGGCAGGCGCCCGTGGGCATCTCCGACGTCTG 821

    244 rLeuPheMetValGlyAspPheLeuLeuAlaValMetGlyPheAlaThrI 261
        :|||   |||:::   ::::::::::::::::::      :::|||   :
    822 GCTCACCATGCTC...AGCATGATCGTGGGTGCCACCTGCTACGCCATGT 868

    261 leMetGlySerMetSerSerValIleTyrAsnMetAsnThrAlaAspAla 277
        :::::|||      ::::::::::|||   ::::::::::::::::
    869 TCATTGGCCACGCCACTGCCCTCATCCAGTCCCTGGACTCCTCCCGGCGC 918

    278 AlaPheTyrProAspHisAlaLeuValLysLysTyrMetLysLeuGlnHi 294
           :::      :::      |||:::::::||||||:::::::::
    919 CAGTACCAGGAAAAGTACAAGCAGGTGGAGCAGTACATGTCCTTTCACAA 968

    294 sValAsnArgLysLeuGluArgArgValIleAspTrpTyrGlnHisLeuG 311
         :::            :::::::|||:::   |||:::|||:::|||   :
    969 GCTCCCGCCCGACACCCGGCAGCGCATCCACGACTACTACGAGCAC...C 1015

    311 lnIleAsnLysLysMetThrAsnGluValAlaIleLeuGlnHisLeuPro 327
        ::   :::   ||||||   :::|||   :::   |||   :::|||
   1016 GCTACCAGGGCAAGATGTTCGACGAGGAGAGCACCCTGGGCGAGCTAAGC 1065

    328 GluArgLeuArgAlaGluValAlaValSerValHisLeuSerThrLeuSe 344
        |||   ||||||   |||:::               :::   :::::
   1066 GAGCCCCTGCGGGAGGAGATCATCAACTTTAACTGTCGGAAGCTGGTGGC 1115

    344 rArgValGlnIlePheGlnAsnCysGluAlaSerLeuLeuGluGluLeuV 361
        :   :::   :::|||   |||:::::::   :::::::::   :::::::
   1116 CTCCATGCCACTGTTTGCCAATGCGGACCCCAACTTCGTGACGTCCATGC 1165

    361 alLeuLysLeuGlnProGlnThrTyrSerProGlyGluTyrValCysArg 377
        ::   ||||||:::   :::::::::::::::||||||:::|||:::   |||
   1166 TGACCAAGCTGCGTTTCGAGGTCTTCCAGCCTGGGGACTACATCATCCGG 1215

    378 LysGlyAspIleGlyGlnGluMetTyrIleIleArgGluGlyGlnLeuAl 394
        :::|||   ||||||:::::::||||||:::|||::::::|||   :::::
   1216 GAAGGCACCATTGGCAAGAAGATGTACTTCATCCAGCATGGCGTGGTCAG 1265

    394 aValVal 396
        :|||:::
   1266 CGTGCTC 1272

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-956-242-1
```

```
seq_documentation_block:
; Sequence 1, Application US/08956242C
; Patent No. 5986081
; GENERAL INFORMATION:
;  APPLICANT: Ganetzky, Barry S.
;  APPLICANT: Titus, Steven A.
;  TITLE OF INVENTION: Polynucleotides Encoding Herg-3
;  FILE REFERENCE: 960296.94550
;  CURRENT APPLICATION NUMBER: US/08/956,242C
;  CURRENT FILING DATE: 1997-10-22
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 3141
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (248)..(2128)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (3)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (12)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1568)..(1872)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (3126)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (3134)
;   OTHER INFORMATION: Unidentified at time of filing
US-08-956-242-1


alignment_scores:
            Quality:  231.00            Length:    349
              Ratio:  1.235               Gaps:     13
 Percent Similarity:  53.582   Percent Identity: 22.063

alignment_block:
 US-09-927-267-1 x US-08-956-242-1   ..

 Align seg 1/1  to: US-08-956-242-1  from: 1  to: 3141

      14 ProProAlaProSerLysAlaArg.LysLeuLeuProValLeuAspProS 30
         ||||||:::|||      :::|||    :::|||||||          :
     535 CCCCCAGGTCCTGTCCCTGGGCGCGGATGTGCTGCCGGAGTACAAGCTGC 584

      30 erGlyAspTyrTyrTyrTrpTrp......LeuAsnThrMetValPhePro 44
         :::::          :::   |||      |||:::          |||
     585 AGGCGCCGCGCATCCACCGCTGGACCATCCTGCACTACAGCCCCTTCAAG 634

      45 ValMetTyrAsn.........LeuIleIleLeuValCysArgAlaCysPh 58
         ::::::::::::         |||:::::::::::::      |||   ||
     635 GCCGTGTGGGACTGGCTCATCCTGCTGCTGGTCATCTACACGGCTGTCTT 684

      58 eProAspLeuGlnHisGlyTyrLeuValAla..................... 68
         |         :::   ::::::|||::::::
     685 CACGCCCTACTCAGCCGCCTTCCTGCTCAGCGATCAGGACGAATCACGGC 734

      69 ............................TrpLeuValLeuAspTyrThr 75
```

```
                                                 |||:::|||
    735 GTGGGGCCTGCAGCTATACCTGCAGTCCCCTCACTGTGGTGGATCTCATC 784

     76 SerAspLeuLeuTyrLeuLeuAspMetValValArgPheHisThrGlyPh 92
        |||:::::::::::::::|||:::|||::::::|||:::|||   ::
    785 GTGGACATCATGTTCGTCGTGGACATCGTCATCAACTTCCGCACCACCTA 834

     92 eLeuGluGlnGlyIleLeuValValAsp...LysGlyArgIleSerSerA 108
        :::::::   :::      ||||||:::         ||||||:::   :
    835 TGTCAACACCAATGATGAGGTGGTCAGCCACCCCCGCCGCATCGCCGTCC 884

    108 rgTyrValArgThrTrpSerPhePheLeuAspLeuAlaSerLeuMetPro 124
        ::|||   :::   |||   |||::::::|||:::::::::   :::|||
    885 ACTACTTCAAGGGCTGG...TTCCTCATTGACATGGTGGCCGCCATCCCT 931

    125 ThrAspValValTyrValArgLeuGlyProHisThrProThr........ 138
        |||::::::      |||   |||            |||
    932 TTCGACCTCCTGATCTTCCGCACTGGCTCCGATGAAACCACAACCCTGAT 981

    139 .......LeuArgLeuAsnArgPheLeuArgAlaProArgLeuPheGluA 153
              |||:::      |||:::||||||      |||:::   :::
    982 TGGGCTATTGAAGACAGCGCGGCTGCTGCGGCTGGTGCGCGTAGCACGGA 1031

    153 laPheAspArgThrGluThrArgThrAlaTyrProAsnAlaPheArgIle 169
          :::||||||   :::         |||                 :::
   1032 AGCTGGACCGCTACTCTGAGTATGGGGCG.............GCTGTG 1066

    170 AlaLysLeuMetLeuTyrIlePheValValIle...HisTrpAsnSerCy 185
           |||::::::      |||::::::|||   ||||||   :::||
   1067 CTCTTCTTGCTCATGTGCACCTTCGCGCTCATAGCGCACTGGCTGGCCTG 1116

    185 sLeuTyrPheAlaLeu.............................. 190
        |::::::::::|||:::
   1117 CATCTGGTACGCCATCGGCAATGTGGAGCGGCCCTACCTAGAACACAAGA 1166

    191 ................................SerArgTyrLeuGly 195
                                       :::||||||   |||
   1167 TCGGGTGGCTGGACAGCCTGGCTGTGCAGCTTGGCAAGCGCTACAACGGC 1216

    196 PheGlyArgAspAlaTrpValTyrProAspProAlaGlnProGlyPheGl 212
                              |||||||||:::
   1217 ......................AGCGACCCAGCCTCGGGC......CC 1236

    212 uArgLeuArgArgGlnTyrLeuTyrSerPheTyrPheSerThrLeuIleL 229
        ::::::   :::|||:::   ::::::||||||:::         |
   1237 CTCGGTGCAGGACAAGTATGTCACAGCCCTCTACTTCACCTTCAGCAGCC 1286

    229 euThrThrValGly...AspThrProProProAlaArgGluGluGluTyr 244
        |||||:::||||||      :::      |||      :::::|||
   1287 TCACCAGCGTGGGCTTCGGCAATGTCTCGCCCAACACCAACTCCGAGAAG 1336

    245 LeuPheMetValGlyAspPheLeuLeuAlaValMetGlyPheAlaThrIl 261
        :::|||   :::      :::|||:::::   :::   :::|||:::||
   1337 GTCTTCTCCATCTGCGTCATGCTCATCGGCTCCCTGATGTACGCCAGCAT 1386

    261 eMetGlySerMetSerSerValIleTyrAsnMetAsnThrAlaAspAlaA 278
        |:::|||::::::|||:::::|||   ::::::   ::::::   |||
   1387 CTTCGGGAACGTGTCCGCGATCATCCAGCGCCTGTACTCGGGAACTGCCA 1436

    278 laPheTyrProAspHisAlaLeuValLysLysTyrMetLysLeuGlnHis 294
          :::::   :::         ||||||:::::::::::::::::::::
   1437 GGTACCACATGCAGATGCTGCGAGTAAAAGAGTTCATTCGCTTTCACCAA 1486

    295 ValAsnArgLysLeuGluArgArgValIleAspTrpTyrGlnHis 309
        :::   :::   |||::::::|||:::   :::::::::||||||
   1487 ATCCCCAACCCTCTGAGGCAACGTCTTGAAGAATATTTCCAGCAC 1531

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-351-215-1

seq_documentation_block:
; Sequence 1, Application US/09351215
; Patent No. 6087488
; GENERAL INFORMATION:
;  APPLICANT: Ganetzky, Barry S.
;  APPLICANT: Titus, Steven A.
;  TITLE OF INVENTION: Polynucleotides Encoding Herg-3
;  FILE REFERENCE: 960296.94550
;  CURRENT APPLICATION NUMBER: US/09/351,215
;  CURRENT FILING DATE: 1999-07-12
;  EARLIER APPLICATION NUMBER: 08/956,242
;  EARLIER FILING DATE: 1997-10-22
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 3141
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (248)..(2128)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (3)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (12)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1568)..(1872)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (3126)
;   OTHER INFORMATION: Unidentified at time of filing
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (3134)
;   OTHER INFORMATION: Unidentified at time of filing
US-09-351-215-1


alignment_scores:
            Quality:  231.00            Length:    349
              Ratio:  1.235               Gaps:     13
 Percent Similarity:  53.582   Percent Identity: 22.063

alignment_block:
 US-09-927-267-1 x US-09-351-215-1   ..

 Align seg 1/1  to: US-09-351-215-1  from: 1  to: 3141

      14 ProProAlaProSerLysAlaArg.LysLeuLeuProValLeuAspProS 30
         ||||||:::|||      :::|||    :::||||||          :
     535 CCCCCAGGTCCTGTCCCTGGGCGCGGATGTGCTGCCGGAGTACAAGCTGC 584

      30 erGlyAspTyrTyrTyrTrpTrp......LeuAsnThrMetValPhePro 44
         :::::          :::   |||     |||::          |||
     585 AGGCGCCGCGCATCCACCGCTGGACCATCCTGCACTACAGCCCCTTCAAG 634

      45 ValMetTyrAsn.........LeuIleIleLeuValCysArgAlaCysPh 58
         ::::::::::::          |||::::::::::::      |||   ||
     635 GCCGTGTGGGACTGGCTCATCCTGCTGCTGGTCATCTACACGGCTGTCTT 684

      58 eProAspLeuGlnHisGlyTyrLeuValAla.................... 68
         |        :::   ::::::|||::::::
     685 CACGCCCTACTCAGCCGCCTTCCTGCTCAGCGATCAGGACGAATCACGGC 734

      69 ............................TrpLeuValLeuAspTyrThr 75
                                            |||:::|||
     735 GTGGGGCCTGCAGCTATACCTGCAGTCCCCTCACTGTGGTGGATCTCATC 784
```

```
     76 SerAspLeuLeuTyrLeuLeuAspMetValValArgPheHisThrGlyPh 92
        |||:::::::::::::::|||:::|||:::::::|||:::|||   ::
    785 GTGGACATCATGTTCGTCGTGGACATCGTCATCAACTTCCGCACCACCTA 834

     92 eLeuGluGlnGlyIleLeuValValAsp...LysGlyArgIleSerSerA 108
        :::::::   :::      ||||||:::         |||||||:::   :
    835 TGTCAACACCAATGATGAGGTGGTCAGCCACCCCCGCCGCATCGCCGTCC 884

    108 rgTyrValArgThrTrpSerPhePheLeuAspLeuAlaSerLeuMetPro 124
        ::|||   :::   |||   |||::::::|||:::::::::   :::|||
    885 ACTACTTCAAGGGCTGG...TTCCTCATTGACATGGTGGCCGCCATCCCT 931

    125 ThrAspValValTyrValArgLeuGlyProHisThrProThr........ 138
        |||::::::      |||   |||             |||
    932 TTCGACCTCCTGATCTTCCGCACTGGCTCCGATGAAACCACAACCCTGAT 981

    139 .......LeuArgLeuAsnArgPheLeuArgAlaProArgLeuPheGluA 153
              |||:::      |||:::||||||      |||:::   :::
    982 TGGGCTATTGAAGACAGCGCGGCTGCTGCGGCTGGTGCGCGTAGCACGGA 1031

    153 laPheAspArgThrGluThrArgThrAlaTyrProAsnAlaPheArgIle 169
          :::||||||   :::         |||                  :::
   1032 AGCTGGACCGCTACTCTGAGTATGGGGCG.............GCTGTG 1066

    170 AlaLysLeuMetLeuTyrIlePheValValIle...HisTrpAsnSerCy 185
            |||::::::      |||::::::|||   ||||||   :::||
   1067 CTCTTCTTGCTCATGTGCACCTTCGCGCTCATAGCGCACTGGCTGGCCTG 1116

    185 sLeuTyrPheAlaLeu.......................... 190
        |:::::::::|||:::
   1117 CATCTGGTACGCCATCGGCAATGTGGAGCGGCCCTACCTAGAACACAAGA 1166

    191 ..............................SerArgTyrLeuGly 195
                                     :::|||||   |||
   1167 TCGGGTGGCTGGACAGCCTGGCTGTGCAGCTTGGCAAGCGCTACAACGGC 1216

    196 PheGlyArgAspAlaTrpValTyrProAspProAlaGlnProGlyPheGl 212
                           |||||||||:::
   1217 ......................AGCGACCCAGCCTCGGGC......CC 1236

    212 uArgLeuArgArgGlnTyrLeuTyrSerPheTyrPheSerThrLeuIleL 229
        ::::::   :::|||:::   ::::::||||||:::          |
   1237 CTCGGTGCAGGACAAGTATGTCACAGCCCTCTACTTCACCTTCAGCAGCC 1286

    229 euThrThrValGly...AspThrProProProAlaArgGluGluGluTyr 244
        |||||:::||||||      :::      |||      :::::::|||
   1287 TCACCAGCGTGGGCTTCGGCAATGTCTCGCCCAACACCAACTCCGAGAAG 1336

    245 LeuPheMetValGlyAspPheLeuLeuAlaValMetGlyPheAlaThrIl 261
        :::|||   :::      :::|||::::::   :::   :::|||:::||
   1337 GTCTTCTCCATCTGCGTCATGCTCATCGGCTCCCTGATGTACGCCAGCAT 1386

    261 eMetGlySerMetSerSerValIleTyrAsnMetAsnThrAlaAspAlaA 278
        |:::|||::::::|||:::::|||   ::::::   ::::::   |||
   1387 CTTCGGGAACGTGTCCGCGATCATCCAGCGCCTGTACTCGGGAACTGCCA 1436

    278 laPheTyrProAspHisAlaLeuValLysLysTyrMetLysLeuGlnHis 294
          ::::::   :::         ||||||:::::::::::::::::::::
   1437 GGTACCACATGCAGATGCTGCGAGTAAAAGAGTTCATTCGCTTTCACCAA 1486

    295 ValAsnArgLysLeuGluArgArgValIleAspTrpTyrGlnHis 309
        :::   :::   |||:::::::|||:::   :::::::::||||||
   1487 ATCCCCAACCCTCTGAGGCAACGTCTTGAAGAATATTTCCAGCAC 1531

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-795-788-1

seq_documentation_block:
; Sequence 1, Application US/08795788
; Patent No. 5795770
;  GENERAL INFORMATION:
;    APPLICANT:  GABER, RICHARD F.
;    TITLE OF INVENTION:  GENETICALLY ENGINEERED EUKARYOTIC
;    TITLE OF INVENTION:  ORGANISM CAPABLE OF DETECTING THE EXPRESSION OF
;    TITLE OF INVENTION:  HETEROLOGOUS ION CHANNELS AND METHOD TO USE SAME
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TILTON, FALLON, LUNGMUS & CHESTNUT
;      STREET:  100 SOUTH WACKER DRIVE, SUITE 960, HARTFORD
;      STREET:  PLAZA
;      CITY:  CHICAGO
;      STATE:  ILLINOIS
;      COUNTRY:  USA
;      ZIP:  60606-4002
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/795,788
;      FILING DATE:  05-FEB-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/923,094
;      FILING DATE:  31-JUL-1992
;      APPLICATION NUMBER:  US 07/874,846
;      FILING DATE:  27-APR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FENTRESS, SUSAN B.
;      REGISTRATION NUMBER:  31,327
;      REFERENCE/DOCKET NUMBER:  NU-9211CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/456-8000
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2173 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-795-788-1


alignment_scores:
            Quality:  219.50              Length:    504
              Ratio:  0.792                 Gaps:     22
 Percent Similarity:  54.960   Percent Identity: 20.833

alignment_block:
 US-09-927-267-1 x US-08-795-788-1   ..

 Align seg 1/1  to: US-08-795-788-1  from: 1  to: 2173

      18 SerLysAlaArgLysLeuLeuProValLeuAspProSerGlyAspTyrTy 34
         :::|||   ||||||         :::::::::|||   :::      ||
     163 ACTAAGCTCCGCAAA......CACATAATCTCTCCTTTTAATCCACGTTA 206

      34 rTyrTrpTrpLeuAsnThrMetValPheProValMetTyrAsnLeuIleI 51
         |     |||          :::|||:::   |||:::|||:::      |
     207 CAGAGCGTGGGAGATGTGGCTAGTATTACTAGTTATTTACTCAGCTTGGA 256

      51 leLeuValCysArgAlaCysPheProAspLeuGlnHisGlyTyrLeuVal 67
         ||        :::   :::|||        :::
     257 TTTGCCCATTTCAATTTGCTTTCATCACCTATAAAAAAGAC......... 297

      68 AlaTrpLeuValLeuAspTyrThrSerAspLeuLeuTyrLeuLeuAspMe 84
         |||   ::::::::::|||         :::   ::::::   :::|||::
     298 GCGATTTTCATCATCGACAACATTGTTAATGGCTTCTTCGCCATTGATAT 347

      84 tValValArgPheHisThrGlyPheLeuGluGlnGly......IleLeuV 99
         :::::::   |||   ::::::::::|||::::::         :::||||
     348 TATTCTCACCTTCTTCGTCGCTTATCTCGATAGCCACTCCTATCTTCTAG 397

      99 alValAspLysGlyArgIleSerSerArgTyrValArgThrTrpSerPhe 115
         ||   :::      :::|||:::   |||||||:::   ||||||   |||
```

```
 398 TTGACAGTCCTAAGAAAATAGCAATAAGGTACCTTTCGACGTGG...TTC 444

 116 PheLeuAspLeuAlaSerLeuMetProThrAspValValTyrVal..... 130
     :::|||:::::|||      |||   :::   :::   :::
 445 GCTTTCGATGTTTGTTCCACAGCACCATTTCAGCCACTAAGCCTCTTGTT 494

 131 ..............ArgLeuGlyProHisThrProThrLeuArgLeuA 142
                   :::||||||               |||   :::
 495 TAACTACAACGGAAGCGAACTAGGA......TTCAGAATTCTTAGCATGC 538

 142 snArgPheLeuArgAlaProArgLeuPheGluAlaPheAspArgThrGlu 158
       |||:::   |||      |||:::   :::   |||   |||   |||
 539 TCAGGTTATGGCGTCTCCGGCGAGTTAGCTCGCTATTTGCAAGGCTTGAG 588

 159 ThrArgThrAlaTyrProAsnAlaPhe......ArgIleAlaLysLeuMe 173
                 :::   |||   |||      |||   :::||||||::
 589 AAAGATATCCGTTTC...AACTATTTCTGGATACGTTGCACAAAACTCAT 635

 173 tLeuTyrIlePheValValIleHisTrpAsnSerCysLeuTyrPheAlaL 190
     :        :::   :::||||||      :::|||:::   :::   :
 636 TTCGGTCACTTTGTTCGCTATACATTGTGCTGGATGTTTCAACTACCTGA 685

 190 euSerArgTyrLeuGlyPheGlyArgAspAlaTrp.........ValTyr 203
     :::::                |||   :::|||         ||||||
 686 TTGCAGATAGATATCCTAATCCAAGAAAGACATGGATTGGAGCTGTGTAT 735

 204 ProAspProAlaGlnProGlyPheGluArgLeuArgArgGlnTyrLeuTy 220
     |||:::      :::   :::          |||   ::::::|||:::
 736 CCAAATTTCAAAGAAGCAAGT.........CTATGGAATAGATATGTGAC 776

 220 rSerPheTyrPheSerThrLeuIleLeuThrThrValGly.......... 233
      ::::::|||:::|||         |||||||||:::|||
 777 TGCTCTTTACTGGTCCATTACGACATTAACGACCACGGGATATGGAGATT 826

 234 ........^..AspThrProProProAlaArgGluGluGluTyrLeuPheMet 247
             ::::::|||                :::|||   ||||||
 827 TTCATGCTGAGAACCCA..............AGAGAAATGCTTTTTGAC 861

 248 ValGlyAspPheLeuLeuAlaValMetGlyPhe...AlaThrIleMetGl 263
     :::      |||::::::      :::|||:::   |||   ::::::||
 862 ATT...TTCTTCATGATGTTCAACCTCGGTTTGACAGCTTACCTCATTGG 908

 263 ySerMetSerSerValIleTyrAsnMetAsnThrAlaAspAlaAlaPheT 280
     |:::|||:::::::::::::   :::   ::::::         :::|||
 909 AAATATGACCAACCTCGTCGTTCATTGGACTAGCCGAACCAGAACCTTTA 958

 280 yrProAspHisAlaLeuValLysLysTyrMetLysLeuGlnHisValAsn 296
         :::          ::::::::::::   :::   :::::::::
 959 GGGATTCAGTGAGAGCTGCTTCAGAGTTTGCTTCAAGAAATCAACTCCCA 1008

 297 ArgLysLeuGluArgArgValIleAspTrpTyrGlnHisLeuGlnIleAs 313
     :::   ::::::   ::::::::::::   :::      |||:::::::::
1009 CATGACATACAAGATCAAATGTTATCA...CACATTTGCTTAAAGTTCAA 1055

 313 nLysLysMetThrAsnGluValAlaIleLeuGlnHisLeuProGluArgL 330
     :   :::      ::::::         |||:::::||||||:::   :
1056 AACAGAGGGCTTGAAACAACAAGAGACCTTGAACAATCTGCCAAAAGCAA 1105

 330 euArgAlaGluValAlaValSerValHisLeuSerThrLeuSerArgVal 346
     ::|||:::::::::|||      :::   :::      :::   ::::::
1106 TCCGGTCAAGCATTGCAAACTATTTATTCTTCCCCATTGTTCACAACATT 1155

 347 GlnIlePheGlnAsnCysGluAlaSerLeuLeuGluGluLeuValLeuLy 363
        :::|||||:::   :::   :::::::|||   :::||||||
1156 TACCTCTTTCAAGGAGTTTCTCGTAACTTCCTCTTTCAATTGGTTTCAGA 1205

 363 sLeuGlnProGlnThrTyrSerProGlyGluTyrValCysArgLysGlyA 380
      ::::::   :::   :::   |||   |||   :::      :::::::
1206 TATAGACGCTGAGTATTTCCCACCAAAAGAAGATATAATTCTACAAAACG 1255

 380 spIleGlyGlnGluMetTyrIleIleArgGluGlyGlnLeu...AlaVal 395
     ::      ::::::||||||:::   :::|||   :::      :::
1256 AAGCTCCTACTGATCTTTACATTCTGGTGTCAGGAGCAGTGGACTTCACT 1305
```

```
 396 ValAlaAspAspGlyIleThrGlnTyr.........AlaValLeuGlyAl 409
     |||      ||||||      |||:::         ||||||:::|||
1306 GTCTACGTTGATGGACATGATCAGTTTCAAGGGAAAGCAGTAATTGGAGA 1355

 409 aGlyLeuTyrPheGlyGluIleSerIleIleAsnIleLysGlyAsnMetS 426
               |||||||||:::::::::::      :::
1356 A......ACATTTGGAGAGGTTGGAGTTTTATACTATAGA.......... 1389

 426 erGlyAsnArgArgThrAlaAsnIleLysSerLeuGlyTyrSerAspLeu 442
           :::      ::::::::::::         |||::::::
1390 ........CCACAACCATTCACAGTAAGAACAACCGAGCTATCTCAAATA 1431

 443 PheCysLeuSerLysGluAspLeuArgGluValLeuSerGluTyrProGl 459
     :::   :::|||:::   :::|||   ::::::::::      :::
1432 CTGCGGATAAGCAGAACATCGCTGATGAGTGCGATGCATGCTCAT..... 1476

 459 nAlaGlnThrIleMetGluGluLysGlyArgGluIleLeuLeuLysMetA 476
                         :::   ||||||   |||         ||||
1477 ..............GCTGACGATGGACGAGTCATC.........ATGA 1501

 476 snLysLeuAspValAsnAlaGluAlaAlaGlu...IleAlaLeuGlnGlu 491
     ||:::|||   ::::::   ::::::   :::   ||||||:::::::::
1502 ACAATCTCTTCATGAAACTTAGAGGGCAACAGTCAATAGCAATAGATGAT 1551

 492 AlaThrGluSer 495
     ::::::   |||
1552 TCGAATACTAGT 1563
```

```
seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US93-03942-1

seq_documentation_block:
; Sequence 1, Application PC/TUS9303942
;  GENERAL INFORMATION:
;    APPLICANT:  GABER, RICHARD F.
;    TITLE OF INVENTION:  GENETICALLY ENGINEERED EUKARYOTIC
;    TITLE OF INVENTION:  ORGANISM CAPABLE OF DETECTING THE EXPRESSION
;    TITLE OF INVENTION:  OF HETEROLOGOUS ION CHANNELS AND METHOD TO
;    TITLE OF INVENTION:  USE SAME
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TILTON, FALLON, LUNGMUS & CHESTNUT
;      STREET:  100 SOUTH WACKER DRIVE, SUITE 960,HARTFORD PLAZA
;      CITY:  CHICAGO
;      STATE:  ILLINOIS
;      COUNTRY:  USA
;      ZIP:  60606-4002
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/03942
;      FILING DATE:  19930421
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/874,846
;      FILING DATE:  27-APR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FENTRESS, SUSAN B.
;      REGISTRATION NUMBER:  31,327
;      REFERENCE/DOCKET NUMBER:  NU-9211CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/456-8000
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2173 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
PCT-US93-03942-1
```

```
alignment_scores:
            Quality:  219.50              Length:    504
              Ratio:  0.792                Gaps:     22
 Percent Similarity:  54.960   Percent Identity: 20.833

alignment_block:
 US-09-927-267-1 x PCT-US93-03942-1   ..

 Align seg 1/1  to: PCT-US93-03942-1  from: 1  to: 2173

      18 SerLysAlaArgLysLeuLeuProValLeuAspProSerGlyAspTyrTy 34
         :::|||   ||||||         ::::::::::|||   :::      ||
     163 ACTAAGCTCCGCAAA......CACATAATCTCTCCTTTTAATCCACGTTA 206

      34 rTyrTrpTrpLeuAsnThrMetValProValMetTyrAsnLeuIleI 51
         |      |||         :::|||:::   |||:::|||:::      |
     207 CAGAGCGTGGGAGATGTGGCTAGTATTACTAGTTATTTACTCAGCTTGGA 256

      51 leLeuValCysArgAlaCysProAspLeuGlnHisGlyTyrLeuVal 67
         ||      :::   :::|||         :::
     257 TTTGCCCATTTCAATTTGCTTTCATCACCTATAAAAAAGAC......... 297

      68 AlaTrpLeuValLeuAspTyrThrSerAspLeuLeuTyrLeuLeuAspMe 84
         |||   ::::::::|||         :::   ::::::   :::||||::
     298 GCGATTTTCATCATCGACAACATTGTTAATGGCTTCTTCGCCATTGATAT 347

      84 tValValArgPheHisThrGlyPheLeuGluGlnGly......IleLeuV 99
         :::::::   |||   ::::::::::|||:::::::         :::||||
     348 TATTCTCACCTTCTTCGTCGCTTATCTCGATAGCCACTCCTATCTTCTAG 397

      99 alValAspLysGlyArgIleSerSerArgTyrValArgThrTrpSerPhe 115
         ||   :::      :::|||:::   ||||||:::   ||||||   |||
     398 TTGACAGTCCTAAGAAAATAGCAATAAGGTACCTTTCGACGTGG...TTC 444

     116 PheLeuAspLeuAlaSerLeuMetProThrAspValValTyrVal..... 130
            :::|||::::::|||      |||   :::   :::   :::
     445 GCTTTCGATGTTTGTTCCACAGCACCATTTCAGCCACTAAGCCTCTTGTT 494

     131 ................ArgLeuGlyProHisThrProThrLeuArgLeuA 142
                         :::|||||||                |||   :::
     495 TAACTACAACGGAAGCGAACTAGGA......TTCAGAATTCTTAGCATGC 538

     142 snArgPheLeuArgAlaProArgLeuPheGluAlaPheAspArgThrGlu 158
           |||:::   |||      |||:::   :::   |||   |||   |||
     539 TCAGGTTATGGCGTCTCCGGCGAGTTAGCTCGCTATTTGCAAGGCTTGAG 588

     159 ThrArgThrAlaTyrProAsnAlaPhe......ArgIleAlaLysLeuMe 173
                    :::   |||   |||      |||   :::|||||::
     589 AAAGATATCCGTTTC...AACTATTTCTGGATACGTTGCACAAAACTCAT 635

     173 tLeuTyrIlePheValValIleHisTrpAsnSerCysLeuTyrPheAlaL 190
         :         :::   :::||||||      :::|||:::   :::   :
     636 TTCGGTCACTTTGTTCGCTATACATTGTGCTGGATGTTTCAACTACCTGA 685

     190 euSerArgTyrLeuGlyPheGlyArgAspAlaTrp.........ValTyr 203
         :::::                  |||   :::|||         ||||||
     686 TTGCAGATAGATATCCTAATCCAAGAAAGACATGGATTGGAGCTGTGTAT 735

     204 ProAspProAlaGlnProGlyPheGluArgLeuArgArgGlnTyrLeuTy 220
         |||:::      :::   :::         |||   ::::::|||:::
     736 CCAAATTTCAAAGAAGCAAGT.........CTATGGAATAGATATGTGAC 776

     220 rSerPheTyrPheSerThrLeuIleLeuThrThrValGly.......... 233
          ::::::|||:::|||         |||||||||:::|||
     777 TGCTCTTTACTGGTCCATTACGACATTAACGACCACGGGATATGGAGATT 826

     234 ........AspThrProProProAlaArgGluGluGluTyrLeuPheMet 247
                 ::::::|||                ::::|||   ||||||
     827 TTCATGCTGAGAACCCA..............AGAGAAATGCTTTTTGAC 861

     248 ValGlyAspPheLeuLeuAlaValMetGlyPhe...AlaThrIleMetGl 263
            :::      |||::::::      :::|||:::   |||   ::::::||
     862 ATT...TTCTTCATGATGTTCAACCTCGGTTTGACAGCTTACCTCATTGG 908

     263 ySerMetSerSerValIleTyrAsnMetAsnThrAlaAspAlaAlaPheT 280
         |:::|||:::::::::::::   :::   ::::::         :::|||
     909 AAATATGACCAACCTCGTCGTTCATTGGACTAGCCGAACCAGAACCTTTA 958

     280 yrProAspHisAlaLeuValLysLysTyrMetLysLeuGlnHisValAsn 296
               :::         ::::::::::::   :::   :::::::::
     959 GGGATTCAGTGAGAGCTGCTTCAGAGTTTGCTTCAAGAAATCAACTCCCA 1008

     297 ArgLysLeuGluArgArgValIleAspTrpTyrGlnHisLeuGlnIleAs 313
         :::   ::::::   ::::::::::::   :::      |||:::::::::
    1009 CATGACATACAAGATCAAATGTTATCA...CACATTTGCTTAAAGTTCAA 1055

     313 nLysLysMetThrAsnGluValAlaIleLeuGlnHisLeuProGluArgL 330
         :   :::      ::::::         |||:::::||||||:::   :
    1056 AACAGAGGGCTTGAAACAACAAGAGACCTTGAACAATCTGCCAAAAGCAA 1105

     330 euArgAlaGluValAlaValSerValHisLeuSerThrLeuSerArgVal 346
         ::|||::::::::::|||      :::   :::      :::   ::::::
    1106 TCCGGTCAAGCATTGCAAACTATTTATTCTTCCCCATTGTTCACAACATT 1155

     347 GlnIlePheGlnAsnCysGluAlaSerLeuLeuGluGluLeuValLeuLy 363
            :::||||||:::   :::   :::::::|||   :::||||||
    1156 TACCTCTTTCAAGGAGTTTCTCGTAACTTCCTCTTTCAATTGGTTTCAGA 1205

     363 sLeuGlnProGlnThrTyrSerProGlyGluTyrValCysArgLysGlyA 380
          ::::::   :::   :::   |||   |||   :::      :::::::
    1206 TATAGACGCTGAGTATTTCCCACCAAAAGAAGATATAATTCTACAAAACG 1255

     380 spIleGlyGlnGluMetTyrIleIleArgGluGlyGlnLeu...AlaVal 395
         ::         ::::::||||||:::   :::|||   :::      :::
    1256 AAGCTCCTACTGATCTTTACATTCTGGTGTCAGGAGCAGTGGACTTCACT 1305

     396 ValAlaAspAspGlyIleThrGlnTyr.........AlaValLeuGlyAl 409
         |||      ||||||      |||:::         ||||||:::|||
    1306 GTCTACGTTGATGGACATGATCAGTTTCAAGGGAAAGCAGTAATTGGAGA 1355

     409 aGlyLeuTyrPheGlyGluIleSerIleIleAsnIleLysGlyAsnMetS 426
               |||||||||:::::::::::::      :::
    1356 A......ACATTTGGAGAGGTTGGAGTTTTATACTATAGA.......... 1389

     426 erGlyAsnArgArgThrAlaAsnIleLysSerLeuGlyTyrSerAspLeu 442
                    :::      ::::::::::::         |||::::::
    1390 ........CCACAACCATTCACAGTAAGAACAACCGAGCTATCTCAAATA 1431

     443 PheCysLeuSerLysGluAspLeuArgGluValLeuSerGluTyrProGl 459
         :::   :::|||:::   :::|||   :::::::::      :::
    1432 CTGCGGATAAGCAGAACATCGCTGATGAGTGCGATGCATGCTCAT..... 1476

     459 nAlaGlnThrIleMetGluGluLysGlyArgGluIleLeuLeuLysMetA 476
                         :::   ||||||   |||         ||||
    1477 ................GCTGACGATGGACGAGTCATC.........ATGA 1501

     476 snLysLeuAspValAsnAlaGluAlaAlaGlu...IleAlaLeuGlnGlu 491
         ||:::|||   ::::::   ::::::   :::   ||||||:::::::::
    1502 ACAATCTCTTCATGAAACTTAGAGGGCAACAGTCAATAGCAATAGATGAT 1551

     492 AlaThrGluSer 495
         ::::::   |||
    1552 TCGAATACTAGT 1563

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-2

seq_documentation_block:
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
```

```
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2


alignment_scores:
            Quality:  128.00              Length:    169
              Ratio:  1.293                 Gaps:      6
 Percent Similarity:  58.580   Percent Identity: 25.444

alignment_block:
 US-09-927-267-1 x US-09-103-840A-2   ..

 Align seg 1/1  to: US-09-103-840A-2  from: 1  to: 4403765

     353 GluAlaSerLeuLeuGluGluLeuValLeuLysLeuGlnProGlnThr.. 368
         |||::::::|||||||||   |||   :::      :::|||:::|||
  4168707GAAGTCAACCTGCTGGAGGCTCTGCGGTTT...GCCAGGCCGGACACGCA 4168753

     369 ..........TyrSerProGlyGluTyrValCysArgLysGlyAspIleG 382
                     :::   |||:::|||:::   :::   ||||||:::
  4168754ACAGATTGAGCTGCAAGCAGGCTCGTATTTGTTCCACGCGGGCGATGTGT 4168803

     382 lyGlnGluMetTyrIleIleArgGluGlyGlnLeuAlaValValAlaAsp 398
         :::::   :::|||::::::|||:::|||:::|||   |||:::|||
  4168804CCGATGCGCTCTACGTGGTGCGCAGCGGCCGCCTGCAAGTCCTCGCCGGC 4168853

     399 AspGlyIleThrGlnTyrAlaVal......LeuGlyAlaGlyLeuTyrPh 413
         ||||||      :::   :::|||      ||||||   |||
  4168854GACGGCGCAAAGGACGAAGTGGTGGCCGAGCTGGGCCGTGGTCAGGTGGT 4168903

     413 eGlyGluIleSerIleIleAsnIleLysGlyAsnMetSerGlyAsnArgA 430
          ||||||::                  |||   :::             |
  4168904CGGGGAGCTC.................GGGGTGCTGCTCGATGCGCCGC 4168935

     430 rgThrAlaAsnIleLysSerLeuGlyTyrSerAspLeuPheCysLeuSer 446
         ||:::|||:::::::::::::::     |||:::|||:::   ::::::
  4168936GGTCCGCGTCGGTTCGTGCGGTACGCGACTCGTCCCTGATGCGAGTGACC 4168985

     447 LysGluAspLeuArgGluVal............................. 453
         |||   ::::::   ::::::
  4168986AAGGCCGAATTCGCGAAGATCGCCGATGCCGGGGTGCTTGGGGCGCTGGC 4169035

     454 ..........LeuSerGluTyrProGlnAlaGlnThrIleMetGluGluL 467
                     ::::::      |||:::::::::      ::::::
  4169036GGGGGTACTGGCCAAACGACAGCACCAGACACGCGTGGCCTCTCAGCGGA 4169085

     467 ysGlyArgGluIleLeuLeuLysMetAsnLysLeuAspValAsnAlaGlu 483
               |||:::::::::   :::      :::|||:::||||||
  4169086CAACGCCGGAGGTCGTTGTCGCGGTCGTCGGTGTCGACGCCAATGCACCG 4169135

     484 AlaAlaGluIleAla......LeuGlnGluAlaThrGluSerArgLeuAr 498
         :::|||   :::|||      |||   :::|||   ::::::||||||||
  4169136GTCGCAATGGTGGCCACCGAATTGTGCAGGGCACTGTCGACACGGCTACG 4169185

     498 gGlyLeu 500
         |::::::
  4169186TGCTGTC 4169192

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-067-351-6
```

```
seq_documentation_block:
; Sequence 6, Application US/09067351
; Patent No. 5994081
;  GENERAL INFORMATION:
;    APPLICANT:  Tang, Y. Tom
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Corley, Neil C.
;    APPLICANT:  Baughn, Mariah
;    TITLE OF INVENTION:  HUMAN KERATINS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  INCYTE PHARMACEUTICALS, INC.
;      STREET:  3174 PORTER DRIVE
;      CITY:  PALO ALTO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Word Perfect 6.1 for Windows/MS-DOS 6.2
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/067,351
;      FILING DATE:  Herewith
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CERRONE, MICHAEL C.
;      REGISTRATION NUMBER:  39,132
;      REFERENCE/DOCKET NUMBER:  PF-0511 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 855-0555
;      TELEFAX:  (650) 845-4166
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1623 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  BRSTNOT07
;      CLONE:  2124178
US-09-067-351-6


alignment_scores:
            Quality:  109.00              Length:    366
              Ratio:  0.661                 Gaps:     16
 Percent Similarity:  45.082   Percent Identity: 20.219

alignment_block:
 US-09-927-267-1 x US-09-067-351-6   ..

 Align seg 1/1  to: US-09-067-351-6  from: 1  to: 1623

     263 GlySerMetSerSerValIleTyrAsnMetAsnThrAlaAspAlaAlaPh 279
         |||:::   ::::::::::::   |||:::|||         |||:::::
     261 GGAAATGGGAAGGCCACCATGCAGAATCTCAACGACCGCCTGGCCTCCTA 310

     279 eTyrProAspHisAlaLeuValLysLysTyrMetLysLeuGlnHisValA 296
         :                 ::::::|||           |||:::::::::|
     311 C.................CTGGAGAAGGTTCGCGCCCTGGAGGAGGCCA 342

     296 snArgLysLeuGluArgArgValIleAspTrpTyrGln..... 308
         ||   |||||||||   |||::::::   |||:::|||
     343 ACATGAAGCTGGAAAGCCGCATCCTGAAATGGCACCAGCAGAGAGATCCT 392

     309 .....................................HisLeuGl 311
                                                 ||||||||
     393 GGCAGTAAGAAAGATTATTCCCAGTATGAGGAAAACATCACACACCTGCA 442

     311 n.........IleAsnLysLysMetThrAsnGluValAlaIleLeuGlnH 325
```

```
443 GGAGCAGATAGTGGATGGTAAGATGACCAATGCTCAGATTATTCTT.... 488

325 isLeuProGluArgLeuArg.............................. 331
489 ..CTCATTGACAATGCCAGGATGGCAGTGGATGACTTCAACCTCAAGTAT 536

332 ..........................AlaGluValAlaValSerValHi 339
537 GAAAATGAACACTCCTTTAAGAAAGACTTGGAAATTGAAGTCGAGGGCCT 586

339 sLeuSerThrLeuSerArgValGlnIlePhe..............GlnA 351
587 CCGAAGGACCTTAGACAACCTGACCATTGTCACAACAGACCTAGAACAGG 636

351 snCysGluAlaSerLeuLeuGluGluLeuValLeuLysLeuGlnProGln 367
637 AGGTGGAAGGAATGAGGAAAGAGCTCATTCTCATGAAGAAGCACCATGAG 686

368 Thr..............TyrSerProGlyGluTyr...ValCysArgLy 378
687 CAGGAAATGGAGAAGCATCATGTGCCAAGTGACTTCAATGTCAATGTGAA 736

378 sGlyAspIleGlyGlnGluMetTyrIleIleArgGluGlyGlnLeuAlaV 395
737 GGTGGATACAGGTCCCAGGGAAGATCTGATTAAG................. 770

395 alValAlaAspAspGlyIleThrGlnTyrAlaValLeuGlyAlaGlyLeu 411
771 ..GTCCTGGAGGATATGAGACAAGAATATGAGCTT.............. 803

412 TyrPheGlyGluIleSerIleIleAsnIleLysGlyAsnMetSerGlyAs 428
804 .................ATAATAAAGAAGAAG................. 818

428 nArgArgThrAlaAsnIleLysSerLeuGlyTyrSerAspLeuPheCysL 445
819 ............................CATCGAGACTTGGACACTT 837

445 euSerLysGluAspLeuArgGluValLeuSerGluTyrProGlnAlaGln 461
838 GGTATAAAGAACAGTCTGCAGCCATGTCCCAGGAGGCAGCCAGTCCAGCC 887

462 ThrIleMetGluGluLysGlyArgGluIleLeuLeuLysMetAsnLysLe 478
888 ACTGTGCAGAGCAGACAAGGT..............GACATCCACGAACT 922

478 uAspValAsnAlaGluAlaAlaGluIleAlaLeuGlnGluAlaThrGluS 495
923 GAAGCGCACATTCCAGGCCCTGGAGATTGACCTGCAGACACAGTACAGCA 972

495 erArgLeuArgGlyLeuAspGlnGlnLeuAspAspLeuGlnThrLysPhe 511
973 CGAAA...TCTGCTTTGGAAAACATGTTATCCGAGACCCAGTCTCGGTAC 1019

512 Ala....................ArgLeuLeuAlaGluLeuGluSerSe 521
1020 TCCTGCAAGCTCCAGGACATGCAAGAGATCATCTCCCACTATGAGGAGGA 1069

521 rAlaLeuLysIleAlaTyrArgIleGluArgLeu......GluTrpGlnT 536
1070 ACTGACGCAGCTACGCCATGAACTGGAGCGGCAGAACAATGAATACCAAG 1119

536 hr.........................ArgGluTrpProMetProGlu 543
1120 TGCTGCTGGGCATCAAAACCCACCTGGAGAAGGAAATCACCACGTACCGA 1169

544 AspLeuAlaGluAlaAspAspGluGlyGluProGluGluGlyThrSer 559
1170 CGGCTCCTGGAGGGAGAGAGTGAAGGGACACGGGAAGAATCAAAGTCG 1217
```

THIS PAGE BLANK (USPTO)

```
OM of: US-09-927-267-1 to: EST:*   out_format : pfs

Date: Jun 20, 2002  5:45 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09927267/runat_19062002_090142_25478/app_query.fasta_1.640
 -DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500
 -MINLEN=0 -MAXLEN=2000000000 -USER=US09927267_@CGN1_1_5197
 -NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-927-267-1
Query length: 575
Database: EST:*
Database sequences: 13736207
Database length: -1841457050
Search time (sec): 1772.320000


score_list:
Sequence            Strd Orig    ZScore    EScore Len  ! Documentation  ..
gb_est2:BI830763     +  962.00 1837.72  4.1e-93   793  ! BI830763 603075162F1 NIH_MGC_11
gb_gss:CNS042AH      -  677.00 1283.59  3.0e-62   923  ! AL271250 Tetraodon nigroviridis
gb_gss:CNS03HKC      -  647.00 1227.24  4.1e-59   800  ! AL244389 Tetraodon nigroviridis
gb_est2:BG293093     +  646.50 1225.06  5.5e-59   882  ! BG293093 602389921F1 NIH_MGC_94
gb_gss:CNS04SGS      +  646.00 1226.98  4.3e-59   699  ! AL305173 Tetraodon nigroviridis
gb_gss:CNS02XOL      +  642.00 1217.26  1.5e-58   819  ! AL218622 Tetraodon nigroviridis
gb_gss:AZ503692      -  639.50 1215.85  1.8e-58   621  ! AZ503692 1M0343L05R Mouse 10kb
gb_est2:BF190904     +  634.00 1206.50  5.9e-58   559  ! BF190904 237569 MARC 2PIG Sus s
gb_gss:AQ537904      +  634.00 1205.08  7.1e-58   627  ! AQ537904 RPCI-11-326D9.TJ RPCI-
gb_est1:AW640105     +  623.50 1186.18  8.0e-57   558  ! AW640105 bl91d05.w1 Blackshear/
gb_est2:BF190899     +  621.00 1181.29  1.5e-56   560  ! BF190899 237563 MARC 2PIG Sus s
gb_est1:BB625423     +  589.00 1117.51  5.3e-53   646  ! BB625423 BB625423 RIKEN full-le
gb_est2:BG342539     +  577.50 1089.56  1.9e-51  1019  ! BG342539 602374710F1 NIH_MGC_9
gb_est2:BF937274     +  559.00 1061.00  7.5e-50   567  ! BF937274 fm62h03.y1 Zebrafish a
gb_est2:BE491188     -  521.50  988.06  8.7e-46   580  ! BE491188 db43c03.x1 Blackshear/
gb_est1:AW638930     +  449.00  849.44  4.6e-38   499  ! AW638930 bl76e03.w1 Blackshear/
gb_est2:W84514       +  428.50  809.65  7.5e-36   502  ! W84514 zd90g10.r1 Soares_fetal_
gb_est2:W58136       +  404.00  763.56  2.8e-33   449  ! W58136 zd18b02.r1 Soares_fetal_
gb_est2:BJ151852     -  393.00  735.99  9.5e-32   743  ! BJ151852 BJ151852 unpublished c
gb_est2:BG304577     +  383.00  720.06  7.3e-31   563  ! BG304577 fl88b05.y1 Zebrafish a
gb_gss:AQ316849      +  364.00  688.19  4.4e-29   378  ! AQ316849 RPCI11-97I5.TJ RPCI-11
gb_est2:BI733974     +  346.50  645.82  1.0e-26   747  ! BI733974 603355503F1 NIH_MGC_94
gb_est2:W26690       -  343.50  643.41  1.4e-26   568  ! W26690 11e4 Human retina cDNA r
gb_est1:AU137778     +  340.00  634.14  4.5e-26   694  ! AU137778 AU137778 PLACE1 Homo s
gb_est2:BJ123339     +  326.00  610.29  9.5e-25   533  ! BJ123339 BJ123339 unpublished c
gb_est2:BM354832     +  324.00  605.51  1.8e-24   573  ! BM354832 rr17e02.y1 Globodera r
gb_est2:BI706445     +  322.00  602.97  2.4e-24   515  ! BI706445 fq06b04.y1 Zebrafish a
gb_est1:AV605886     -  322.00  602.89  2.5e-24   518  ! AV605886 AV605886 Bos taurus ki
gb_est2:BG473313     +  320.50  594.54  7.2e-24   803  ! BG473313 602515428F1 NIH_MGC_16
gb_est1:AI387006     +  307.50  572.46  1.2e-22   625  ! AI387006 GH17414.5prime GH Dros
gb_gss:CNS025BE      +  305.50  562.16  4.6e-22  1048  ! AL181859 Tetraodon nigroviridi
gb_est2:BI736222     +  292.50  541.33  6.6e-21   738  ! BI736222 603359912F1 NIH_MGC_94
gb_est2:BE287002     +  289.00  532.35  2.1e-20   881  ! BE287002 601097229F1 NCI_CGAP_M
gb_gss:BH121161      -  288.00  539.92  7.9e-21   410  ! BH121161 RPCI-24-393E3.TV RPCI-
gb_gss:CNS01GHZ      +  280.00  517.32  1.4e-19   726  ! AL143160 Anopheles gambiae GSS
gb_gss:CNS04E7T      -  277.50  512.12  2.8e-19   747  ! AL286706 Tetraodon nigroviridis
gb_est1:BB635459     +  277.00  512.78  2.6e-19   655  ! BB635459 BB635459 RIKEN full-le
gb_est2:BI548103     +  274.00  506.18  6.0e-19   698  ! BI548103 603189424F1 NIH_MGC_95
gb_est2:BG974320     +  264.00  485.33  8.7e-18   786  ! BG974320 602844071F1 NIH_CGAP_M
gb_est1:AI035456     +  262.00  488.27  6.0e-18   454  ! AI035456 ub47c01.r1 Soares_mamm
gb_est2:T97907       +  257.00  480.93  1.5e-17   376  ! T97907 ye54c11.r1 Soares fet
gb_htc:AK014722      +  247.50  442.46  2.1e-15  1888  ! AK014722 Mus musculus 0 day
gb_est2:BF725470     +  246.00  454.82  4.4e-16   553  ! BF725470 bx16c03.y1 Human Ir
gb_est1:AL120448     +  238.00  444.89  1.6e-15   353  ! AL120448 DKFZp761I128_r1 761
gb_est1:AL120239     +  238.00  444.11  1.7e-15   376  ! AL120239 DKFZp761H167_r1 761

seq_name: gb_est2:BI830763

seq_documentation_block:
LOCUS       BI830763                793 bp    mRNA    linear   EST 04-OCT-2001
DEFINITION  603075162F1 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5167238 5',
            mRNA sequence.
ACCESSION   BI830763
VERSION     BI830763.1  GI:15942313
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 793)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11415  row: k  column: 15
            High quality sequence stop: 723.
FEATURES             Location/Qualifiers
     source          1. .793
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5167238"
                     /clone_lib="NIH_MGC_119"
                     /tissue_type="medulla"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: EcoRV (destroyed); RNA source normal medulla from
                     anonymous male age 27.  Library  is oligo-dT primed and
                     directionally cloned (EcoRV site is destroyed  upon
                     cloning).  Average insert size 1.3 kb, insert size range
                     0.9-3 kb.  Library is normalized and enriched for
                     full-length clones and  was constructed by C. Gruber
                     (Invitrogen).  Research Genetics  tracking code 013. Note:
                     this is a NIH_MGC Library."
BASE COUNT      160 a    222 c    232 g    179 t
ORIGIN


alignment_scores:
           Quality:  962.00            Length:    234
             Ratio:  4.393               Gaps:     12
 Percent Similarity:  93.590   Percent Identity: 90.598

alignment_block:
 US-09-927-267-1 x BI830763   ..

 Align seg 1/1  to: BI830763  from: 1  to: 793

      21 ArgLysLeuLeuProValLeuAspProSerGlyAspTyrTyrTyrTrpTr 37
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     116 AGGAAGTTGCTGCCTGTCCTGGACCCATCTGGGGATTACTACTACTGGTG 165

      37 pLeuAsnThrMetValPheProValMetTyrAsnLeuIleIleLeuValC 54
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     166 GCTGAACACAATGGTCTTCCCAGTCATGTATAACCTCATCATCCTCGTGT 215
```

```
     54 ysArgAlaCysPheProAspLeuGlnHisGlyTyrLeuValAlaTrpLeu 70
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    216 GCAGAGCCTGCTTCCCCGACTTGCAGCACGGTTATCTGGTGGCCTGGTTG 265

     71 ValLeuAspTyrThrSerAspLeuLeuTyrLeuLeuAspMetValValAr 87
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    266 GTGCTGGACTACACGAGTGACCTGCTATACCTACTAGACATGGTGGTGCG 315

     87 gPheHisThrGlyPheLeuGluGlnGlyIleLeuValValAspLysGlyA 104
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    316 CTTCCACACAGGATTCTTGGAACAGGGCATCCTGGTGGTGGACAAGGGTA 365

    104 rgIleSerSerArgTyrValArgThrTrpSerPhePhe.LeuAspLeuAl 120
        |||||||||||||||||||||||||||||||||| ||||||||||||
    366 GGATCTCGAGTCGCTACGTTCGCACCTGGAGTTTCTTCTTTGGACCTGGC 415

    120 aSerLeuMetProThrAspValValTyrVal.ArgLeuGlyProHisThr 136
        |||||||||||||||||||||||||||||| ||||||||||||:::|||
    416 TTCCCTGATGCCCACAGATGTGGTCTACGTTGCGGCTGGGCCCGTACACA 465

    137 Pro.ThrLeuArgLeuAsnArgPheLeuArgAlaProArgLeuPheGluA 153
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    466 ACACACCCTGAGGCTGAACCGCTTTCTCCGCGCGCCCCGCCTCTTCGAGG 515

    153 laPheAspArgThrGluThrArgThrAlaTyrProAsnAlaPheArgIle 169
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    516 CCTTCGACCGCACAGAGACCCGCACAGCTTACCCAAATGCCTTTCGCATT 565

    170 AlaLysLeuMetLeuTyrIlePheValValIleHisTrp.AsnSerCysL 186
        |||||||||||||||||||   |||||||||||||| :::|||||||
    566 GCCAAGCTGATGCTTTACATTAT.GTCGTCATCCATTGGAAGCAGCTGCC 614

    186 eu.TyrPheAlaLeuSerArgTyrLeuGlyPheGlyArg.AspAlaTrpV 202
        || ||||||||||||||||||||||||||||||||| ||||||||||
    615 TATTACTTTGCCCTATCCCGGTACCTGGGCTTCGGGCGTTGACGCATGGG 664

    202 alTyrProAsp.ProAlaGlnPro.GlyPheGluArgLeuArg.ArgGln 217
        |||||||||::: |||:::|||||| |||||||||||||||||| ||||||
    665 TGTACCCGGAACCCGACGCAGCCTTGGCTTTGAGCGACTGCGGGCGCCAG 714

    218 TyrLeuTyrSerPheTyrPheSerThrLeu.IleLeuThrThrValGly. 233
        |||||||||||||||||||||||||||||| ||||||||||||||||||
    715 TACCTCTATAGCTTTTACTTCTCCACGCTGAATACTGACTACAGTGGGCG 764

    234 AspThr.ProProProAlaArgGluGlu 242
        |||||| |||      :::|||:::|||
    765 GATACAACCGACGGCAAGCAGGAAAGAA 792

seq_name: gb_gss:CNS042AH

seq_documentation_block:
LOCUS       CNS042AH                923 bp    DNA     linear   GSS 18-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            077A21 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL271250
VERSION     AL271250.1  GI:7993216
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 923)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 923)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 923)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .923
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="077A21"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG077AA11LP1-end : T7"
BASE COUNT      256 a    208 c    264 g    194 t      1 others
ORIGIN


alignment_scores:
            Quality:  677.00             Length:    295
              Ratio:  3.239                Gaps:      4
 Percent Similarity:  70.847   Percent Identity: 43.390

alignment_block:
 US-09-927-267-1 x CNS042AH/rev   ..

 Align seg 1/1  to reverse of: CNS042AH  from: 1  to: 923

     36 TrpTrpLeuAsnThrMetValProValMetTyrAsnLeuIleIleLe 52
        ||||||   :::   ::::::   |||:::      :::|||::::::||
    918 TGGTGGAGGAGCCGATTGACC...CCCATA......ACGCTGGTGCTGCT 878

     52 uValCysArgAlaCysProAspLeuGlnHisGlyTyrLeuValAlaT 69
        |:::   ||||||||||||   :::||||||||:::::::::|||   |
    877 CCTCCTCAGGGCTTGCTTCGAAGAGCTGCAGCATAACTTTATCGTCTACT 828

     69 rpLeuValLeuAspTyrThrSerAspLeuLeuTyrLeuLeuAspMetVal 85
        ||::::::|||||:::||||||||||||:::|||:::   ||||||
    827 GGTTTCTCCTTGACTTCACATCGGACCTGATCTACGTGGCGGACATGTTT 778

     86 ValArgPheHisThr..................................... 90
        ||||||   :::|||
    777 GTGAGGACCAGAACCGGTACCGCGTCACCGTGTCAGCCTGTTCTCAGCGT 728

     90 ..................................................... 90

    727 GCAAACCAATGTCATCGTGAGGGAGGGTCCACTTGAGTGTTTTCTGTCTC 678

     91 ...........GlyPheLeuGluGlnGlyIleLeuValValAspLysGly 103
                  |||:::||||||||||||:::||||||   |||:::
    677 TTCCGAGTGCAGGTTACCTGGAACAAGGCCTGCTGGTGAAAGACGAGAAG 628

    104 ArgIleSerSerArgTyrValArgThrTrpSerPhePheLeuAspLeuAl 120
          :::   :::   ||||||   :::::::::|||   ||||||:::::
    627 CTCCTCAGGGACGGCTACGTTGCCAGCTTTCAGTTTCGTCTTGATGTCGT 578

    120 aSerLeuMetProThrAspValValTyrValArgLeuGlyProHisThrP 137
        :|||::::::|||||||:::::|||:::   :::|||         |
    577 TTCCATGCTTCCCACTGACATCTTGTACTTGTGCTTTGGCCTCGACTACC 528

    137 roThrLeuArgLeuAsnArgPheLeuArgAlaProArgLeuPheGluAla 153
        ||   :::|||:::|||::::::||||||      |||:::::::|||
    527 CAGAGATCCGAATCAACAAGCTTTTGAGAATCGGCCGGATGATGGAGTTC 478

    154 PheAspArgThrGluThrArgThrAlaTyrProAsnAlaPheArgIleAl 170
        |||   ||||||||||||:::|||   |||||||||   ||||||||||
    477 TTCACCAGGACCGAGACTAAAACCAACTACCCCAACATCTTCCGCATTGC 428
```

```
     170 aLysLeuMetLeuTyrIlePheValValIleHisTrpAsnSerCysLeuT 187
         |:::|||::::::||||||:::::::::|||||||||||||:::|||:::|
     427 AAACCTGATTATGTACATCCTCATTATCATCCACTGGAACGCTTGCTTTT 378

     187 yrPheAlaLeuSerArgTyrLeuGlyPheGlyArgAspAlaTrpValTyr 203
         |||||::::::|||:::   :::|||||||||   |||   |||||||||
     377 ACTTCTCCTTTTCCAAATCTATCGGTTTTGGCGCTGACGAATGGGTCTAC 328

     204 Pro.........AspProAlaGlnProGlyPheGluArgLeuArgArgGl 217
         |||         ||||||   :::|||:::|||   :::      |||::
     327 CCAGCTCTGGACGATCCGGACGAGCCTGCTTTCGGGGAGGCCATGAGGAA 278

     217 nTyrLeuTyrSerPheTyrPheSerThrLeuIleLeuThrThrValGlyA 234
         :|||   :::|||:::|||:::|||||||||   |||||||||:::|||:
     277 GTATTCTTTCAGCCTGTACTGGTCTACCCTGACCCTCACCACCATTGGAG 228

     234 spThrProProProAlaArgGluGluGluTyrLeuPheMetValGlyAsp 250
         ::|||||||||||||||   ::::::|||:::||||||   |||   |||
     227 AAACCCCCCCACCGGCCCTGGACTCGGAATTCCTCTTTCATGTGATTGAT 178

     251 PheLeuLeuAlaValMetGlyPheAlaThrIleMetGlySerMetSerSe 267
         ||||||:::::|||:::   |||||||||||:::|||:::::::::::
     177 TTCCTAGTTGGCGTCCTGATTTTTGCCACCATTGTAGGAAACATCGCCAC 128

     267 rValIleTyrAsnMetAsnThrAlaAspAlaAlaPheTyrProAspHisA 284
         ::::|||   |||||||||:::|||:::|||   |||
     127 CATGATCTCAAACATGAACGCCGCCCAGGCTCAGTTCCAGGCTCGGATCG 78

     284 laLeuValLysLysTyrMetLysLeuGlnHisVal 295
           :::|||   ||||||:::::::|||:::::
      77 ACAACATGAAACACTACATGCAGGTACAGCAGCTT 43

seq_name: gb_gss:CNS03HKC

seq_documentation_block:
LOCUS       CNS03HKC                800 bp    DNA     linear   GSS 17-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            027I05 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL244389
VERSION     AL244389.1  GI:7965401
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 800)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 800)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 800)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .800
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="027I05"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG027AE03SP1-end :
                     PUC-Ori"
BASE COUNT      173 a    228 c    195 g    199 t      5 others
ORIGIN


alignment_scores:
           Quality:  647.00            Length:    256
             Ratio:  3.387               Gaps:      3
 Percent Similarity:  74.609   Percent Identity: 50.391

alignment_block:
 US-09-927-267-1 x CNS03HKC/rev   ..

 Align seg 1/1  to reverse of: CNS03HKC  from: 1  to: 800

     275 AlaAspAlaAlaPheTyrProAspHisAlaLeuValLysLysTyrMetLy 291
         |||:::|||   |||                    :::|||   ||||||::
     798 GCCCAGGCTCAGTTCCAGGCTCGRATCGACAACATCAAACACTACATGCA 749

     291 sLeuGlnHis..................................... 294
         ::::|||:::
     748 GGTACAGCAGCTTTAGACCGCATTCTCTTTTCTCATTGCTTTTTTTTCAT 699

     295 .....:.................................ValAsnArgLys 298
                                              . |||::::::
     698 ATTTTGAATCTTTTTCTTCGGGTCCACAGGCCCGGAAGGTCAGCAAGGAC 649

     299 LeuGluArgArgValIleAspTrpTyrGlnHisLeuGlnIleAsnLysLy 315
         ||||||   |||||||||   |||::::::::::|||      |||   ||
     648 CTCGAGTTGCGAGTCATCAAGTGGTTTGACTATTTGTGGAACAACGGAAA 599

     315 sMetThrAsnGluValAlaIleLeuGlnHisLeuProGluArgLeuArgA 332
         |      :::|||      :::|||::::::||||||:::::::|||:::|
     598 AGCCCAGGATGAGAGAGAGGTGCTGAGGTACCTTCCGGACAAGCTCAAGG 549

     332 laGluValAlaValSerValHisLeuSerThrLeuSerArgValGlnIle 348
         |||||:::|||::::::||||||::::::||||||::::::|||:::|||
     548 CGGAGATCGCCATCCAGGTCCACATGGATACTCTGAAGAAAGTCCGTATT 499

     349 PheGlnAsnCysGluAlaSerLeuLeuGluGluLeuValLeuLysLeuGl 365
         |||::::::||||||||||:::||||||   |||||||||||||||||::
     498 TTTTCCGACTGCGAGGCGGGCCTGTTGATCGAGCTGGTGCTCAAGCTGCG 449

     365 nProGlnThrTyrSerProGlyGluTyrValCysArgLysGlyAspIleG 382
         :||||||::::::||||||||||:::|||:::|||:::||||||||||||
     448 GCCGCAGGTGTTCAGCCCCGGGGACTACATCTGCAAGAAGGGCGACATTG 399

     382 lyGlnGluMetTyrIleIleArgGluGlyGlnLeuAlaValValAlaAsp 398
         ||:::|||||||||||||||||::::::|||:::|||||||||||||||||
     398 GCCGTGAGATGTACATCATCAAGGACGGAAAGCTGGCGGTTGTCGCCGAT 349

     399 AspGlyIleThrGlnTyrAlaValLeuGlyAlaGlyLeuTyrPheGlyGl 415
         ||||||:::||||||::::::||||||||||:::|||   ||||||||||
     348 GATGGTGTCACGCAATTTGTGGTGCTGGGAAGTGGCAGTTATTTCGGCGA 299

     415 uIleSerIleIleAsnIleLysGlyAsnMetSerGlyAsnArgArgThrA 432
         ||||||||||:::|||||||||||||:::   :::|||||||||||||||
     298 GATCAGCATCCTTAACATCAAAGGCAGCAAGGCCGGGAACCGTCGAACGG 249

     432 laAsnIleLysSerLeuGlyTyrSerAspLeuPheCysLeuSerLysGlu 448
         ||||||||:::|||:::||||||||||||||||||||||||||||||:::
     248 CGAACATCCGCAGCATCGGTTACTCGGACCTCTTCTGCCTCTCCAAGGAC 199

     449 AspLeuArgGluValLeuSerGluTyrProGlnAlaGlnThrIleMetGl 465
         ||||||   |||   |||   |||||||||:::|||:::::::::::::||
     198 GACCTGATGGAGTCTTTGTTGGAGTATCCGGATGCCAAAGCCATGCTGGA 149

     465 uGluLysGlyArgGluIleLeuLeuLysMetAsnLysLeuAsp....... 479
```

```
     148 GGAGAAGGGCCGGCAAATCCTGATGAAGGACGGTCTGTTGGACCTGGACC 99

     480 ........ValAsnAlaGluAlaAlaGluIleAlaLeuGlnGluAlaThr 493
                 ::::::    ||||||   |||:::        |||:::
      98 CGGCGAACATCAAGCCCGAAGCCAAAGAGCTG.........GARRARAAG 58

     494 GluSerArgLeuArgGly 499
         :::::::|||   |||
      57 GTCAACAAGCTCTACGGC 40

seq_name: gb_est2:BG293093

seq_documentation_block:
LOCUS       BG293093                882 bp    mRNA    linear   EST 21-FEB-2001
DEFINITION  602389921F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4501732 5',
            mRNA sequence.
ACCESSION   BG293093
VERSION     BG293093.1  GI:13052518
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 882)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10369  row: j  column: 05
            High quality sequence stop: 685.
FEATURES             Location/Qualifiers
     source          1. .882
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4501732"
                     /clone_lib="NIH_MGC_94"
                     /tissue_type="retina"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 3.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      261 a    178 c    203 g    240 t
ORIGIN


alignment_scores:
            Quality:  646.50            Length:    255
              Ratio:  3.123               Gaps:      7
 Percent Similarity:  81.176   Percent Identity: 50.980

alignment_block:
 US-09-927-267-1 x BG293093   ..

 Align seg 1/1  to: BG293093  from: 1  to: 882

     156 ArgThrGluThrArgThrAlaTyrProAsnAlaPheArgIleAlaLysLe 172
         ||||||||||||||||||   |||||||||   |||||||||:::::::||
       2 AGGACGGAAACAAGGACCAACTACCCAAACATCTTTAGGATTTCGAACCT 51

     172 uMetLeuTyrIlePheValValIleHisTrpAsnSerCysLeuTyrPheA 189
         |:::::::||||||   ::::::|||||||||||||||:::|||:::|||:::::
      52 TGTGATGTACATCGTCATCATCATCCACTGGAACGCTTGTGTGTATTATT 101

     189 laLeuSerArg..TyrLeuGlyPheGlyArgAspAlaTrpValTyrProA 205
         :::::|||:::  |||   |||||||||:::|||:::|||||||||||||
     102 CCATCTCAAAAGCTAT...GGATTTGGGAATGACACATGGGTCTACCCTG 148

     205 spProAlaGlnProGlyPheGluArgLeuArgArgGlnTyrLeuTyrSer 221
         ||       :::|||   |||   ||||||   |||:::|||:::||||||
     149 ATGTTAATGATCCTGAATTCGGCCGTTTAGCTAGAAAATATGTCTACAGC 198

     222 PheTyrPheSerThrLeuIleLeuThrThrValGlyAspThrProProPr 238
         :::|||:::|||||||||   ||||||||||:::|||:::||||||||||
     199 CTTTATTGGTCTACTTTGACTTTGACAACCATTGGTGAAACTCCACCTCC 248

     238 oAlaArgGluGluGluTyrLeuPheMetValGlyAspPheLeuLeuAlaV 255
         |:::   ::::::||||||:::|||:::|||   |||||||||:::::::|
     249 CGTGCTGGATTCTGAGTACATCTTTGTGGTGGTAGACTTCTTAATTGGAG 298

     255 alMetGlyPheAlaThrIleMetGlySerMetSerSerValIleTyrAsn 271
         ||:::   ||||||||||||:::|||:::::::::|||:::|||   |||
     299 TTTTAATTTTCGCCACCATTGTCGGTAACATAGGTTCCATGATTTCCAAT 348

     272 MetAsnThrAlaAspAlaAlaPheTyrProAspHisAlaLeuValLysLy 288
         ||||||:::|||   |||   |||                  :::|||::
     349 ATGAATGCAGCCCGGGCAGAATTTCAATCAAGAGTTGATGCTATCAAACA 398

     288 sTyrMetLysLeuGlnHisValAsnArgLysLeuGluArgArgValIleA 305
         :||||||:::::::::::::|||:::::::   :::|||:::|||||||||
     399 GTACATGAATTTTCGAAATGTAAGCAAAGATATGGAAAAGAGGGTTATTA 448

     305 spTrpTyrGlnHisLeuGlnIleAsnLysLysMetThrAsnGluValAla 321
           |||:::::::::|||       |||||||||   :::::::|||
     449 AATGGTTTGACTACCTGTGGACCAACAAAAAAACAGTGGATGAGAGAGAA 498

     322 IleLeuGlnHisLeuProGluArgLeuArgAlaGluValAlaValSerVa 338
         :::|||:::::::||||||::::::||||||||||||:::|||::::::||
     499 GTCCTGAGATACCTCCCTGACAAACTAAGGGCAGAAATTGCCATCAACGT 548

     338 lHisLeuSerThrLeuSerArgValGlnIlePheGlnAsnCysGluAlaS 355
         |||||||:::|||||||::::::|||:::||||||   :::||||||||||:
     549 TCATCTAGACACGTTAAAAAAGGTTCGCATCTTTGCTGACTGTGAAGCTG 598

     355 erLeuLeuGluGluLeuValLeuLys...LeuGlnProGlnThrTyrSer 370
         ::|||      ||||||||||||:::      :::|||::::::||||||
     599 GTCTGTGGTGGGAGTTGGTGTTGGAAACTACAAACCCCAGGGTGTATAGT 648

     371 ProGlyGlu...TyrVal.CysArgLysGlyAspIle......GlyGlnG 384
         |||   |||   |||::: ||||||||||||||||||      |||:::
     649 CCCCGGGAGATTTACATTTTGCAGAAAAGGGGACATCCGCGCGGGGAGAA 698

     384 luMetTyrIleIleArgGluGlyGlnLeuAlaValValAlaAspAspGly 400
           :::      |||::::::::::|||   ||||||:::
     699 TGTACCATCATCACAGGAAGGCAAACTTGCCTGTGGTGGGCAAA..... 742

     401 IleThrGlnTyr 404
         :::|||:::|||
     743 TTGACCGATTAC 754

seq_name: gb_gss:CNS04SGS

seq_documentation_block:
LOCUS       CNS04SGS                699 bp    DNA     linear   GSS 24-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            007004 of library H from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL305173
VERSION     AL305173.1  GI:8196132
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
```

```
REFERENCE   1  (bases 1 to 699)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 699)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 699)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .699
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="007004"
                     /clone_lib="H"
                     /note="Genoscope sequence ID : C0BH007BH02XE1-end :
                     PUC-Ori"
BASE COUNT      162 a    216 c    151 g    166 t      4 others
ORIGIN


alignment_scores:
           Quality:  646.00            Length:    229
             Ratio:  3.569               Gaps:      0
 Percent Similarity:  79.039   Percent Identity: 52.402

alignment_block:
 US-09-927-267-1 x CNS04SGS   ..

 Align seg 1/1  to: CNS04SGS  from: 1  to: 699

      91 GlyPheLeuGluGlnGlyIleLeuValValAspLysGlyArgIleSerSe 107
         ||||||||||||||||||:::|||||||   |||:::::::::::::   ::
       9 GGCTTTCTRGAACAAGGAATGCTGGTGAAGGACCAAGCAAAGCTGTGGGC 58

     107 rArgTyrValArgThrTrpSerPhePheLeuAspLeuAlaSerLeuMetP 124
         :      :::|||||||   :::|||   |||||||||   |||:::::::|
      59 CAGCTAMATCCGAACCCACCAGTTCAAACTGGACTTGCTGTCCATCCAGC 108

     124 roThrAspValValTyrValArgLeuGlyProHisThrProThrLeuArg 140
         |||||   ::::::|||:::      |||   |||   |||   ||||||
     109 CCACTGAMCTGGCCTACATCTCCACGGGCATCCACAAGCCGCAGCTCAGG 158

     141 LeuAsnArgPheLeuArgAlaProArgLeuPheGluAlaPheAspArgTh 157
         :::||||||:::||||||   ||||||:::||||||   ||||||||||
     159 TTCAACCGCCTCCTCCGCTTCCCGCGCATGTTCGAATTCTTCGATCGCAC 208

     157 rGluThrArgThrAlaTyrProAsnAlaPheArgIleAlaLysLeuMetL 174
         ||||||||||||   |||||||||   |||||||||::::::|||:::|
     209 CGAGACTCGCACCAACTACCCCAACATCTTCCGTATTTGCAACCTGGTGC 258

     174 euTyrIlePheValValIleHisTrpAsnSerCysLeuTyrPheAlaLeu 190
         ||||||||:::|||:::||||||||||||:::|||:::|||:::|||:::
     259 TCTACATCCTGGTCATCATTCACTGGAACGCCTGCATATATTACGCCATA 308

     191 SerArgTyrLeuGlyPheGlyArgAspAlaTrpValTyrProAspProAl 207
         |||:::   ||||||||||||   |||:::|||||||:::|||:::   ::
     309 TCCAAGTCTTTGGGGTTCGGTTCGGATACTTGGGTTTTCCCAAATATTTC 358

     207 aGlnProGlyPheGluArgLeuArgArgGlnTyrLeuTyrSerPheTyrP 224
         :   |||   ::::::   |||   |||:::   :::|||   :::|||:
     359 CATACCGGAATATTCCTCCCTGACTCGGAGCTAMGTCTACTGCCTCTACT 408

     224 heSerThrLeuIleLeuThrThrValGlyAspThrProProProAlaArg 240
         ::|||||||||   ||||||||||:::|||:::   |||   |||:::|||
     409 GGTCGACTCTTACTCTCACCACCATTGGAGAGATGCCCGCCCCCGTCCGA 458

     241 GluGluGluTyrLeuPheMetValGlyAspPheLeuLeuAlaValMetGl 257
         :::|||||||||||||||||:::|||   ||||||||:::::::|||:::
     459 GACGAAGAGTACCTCTTTGTGGTCTTTGATTTTCTCGTCGGCGTGCTGAT 508

     257 yPheAlaThrIleMetGlySerMetSerSerValIleTyrAsnMetAsnT 274
         ||||||||||||:::|||::::::::::|||:::|||   |||||||||:
     509 CTTCGCCACAATTGTGGGGAACGTCGGCTCCATGATTGCCAACATGAACG 558

     274 hrAlaAspAlaAlaPheTyrProAspHisAlaLeuValLysLysTyrMet 290
         :::::   |||   |||                   :::|||   ||||||
     559 CCACCCGAGCCGAGTTCCAGGCTCGGATCGACGCCATCAAACACTACATG 608

     291 LysLeuGlnHisValAsnArgLysLeuGluArgArgValIleAspTrpTy 307
         ::::::   ||||||:::::::||||||   |||||||||   |||::
     609 CACTTCCGAAAAGTCAACAAAGAGCTGGAAACTCGCGTCATCAAGTGGTT 658

     307 rGlnHisLeuGlnIleAsnLysLysMetThrAsnGlu 319
         ::::::::|||      |||||||||   :::::::|||
     659 CGACTACCTCTGGACCAACAAGAAAGCTGTGGACGAG 695

seq_name: gb_gss:CNS02XOL

seq_documentation_block:
LOCUS       CNS02XOL                819 bp    DNA     linear   GSS 15-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            179G05 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL218622
VERSION     AL218622.1  GI:7877441
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 819)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 819)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 819)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .819
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="179G05"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG179AD03SP1-end :
                     PUC-Ori"
```

```
BASE COUNT      217 a    236 c    180 g    186 t
ORIGIN

alignment_scores:
            Quality:  642.00              Length:    242
              Ratio:  3.309                Gaps:      0
 Percent Similarity:  80.165   Percent Identity: 47.521

alignment_block:
 US-09-927-267-1 x CNS02XOL   ..

 Align seg 1/1  to: CNS02XOL  from: 1  to: 819

      92 PheLeuGluGlnGlyIleLeuValValAspLysGlyArgIleSerSerAr 108
         :::|||||||||||||:::||||||   |||      ::::::   :::::
      92 TACTTGGAGCAAGGCCTTCTGGTGAAAGACGCCAAGAAACTGAGAGATAA 141

     108 gTyrValArgThrTrpSerPhePheLeuAspLeuAlaSerLeuMetProT 125
         :|||      |||   :::|||      |||:::   |||::::::||||
     142 ATACCGAACAACACCCCAGTTCAAGTACGACATGATTTCCATGATACCCA 191

     125 hrAspValValTyrValArgLeuGlyProHisThrProThrLeuArgLeu 141
         |||||:::::::::::::::   |||   ::::::|||   :::|||:::
     192 CCGACCTGCTGTTTCTCAACTACGGCTACAACAATCCGGAGTTCCGCTTC 241

     142 AsnArgPheLeuArgAlaProArgLeuPheGluAlaPheAspArgThrGl 158
         ||||||:::   :::      ||||||||||||   |||:::||||||||
     242 AACCGCCTTTGCAAAATCTCAAGACTTTTTGAATTTTTTGAAAGGACTGA 291

     158 uThrArgThrAlaTyrProAsnAlaPheArgIleAlaLysLeuMetLeuT 175
         ||||||||||:::::||||||   |||||||||:::::|||:::||||
     292 AACTAGAACCAGCTTCCCCAACATGTTCCGTATCAGCAACCTAGTGCTCT 341

     175 yrIlePheValValIleHisTrpAsnSerCysLeuTyrPheAlaLeuSer 191
         |||||:::|||:::|||||||||||||:::|||||||:::||||||||||||
     342 ACATCCTCGTCATCATCCACTGGAACGCCTGCCTGTTCTTCGCCCTCTCC 391

     192 ArgTyrLeuGlyPheGlyArgAspAlaTrpValTyrProAspProAlaGl 208
         :::   :::|||||||||   |||:::||||||||||||:::   :::::
     392 AAAACCATCGGCTTTGGCAGTGACACCTGGGTGTATCCCAACATCAGCCA 441

     208 nProGlyPheGluArgLeuArgArgGlnTyrLeuTyrSerPheTyrPheS 225
         :|||        ||||||   |||:::|||:::||||||:::|||:::|
     442 CCCGGAGCATGGCCGCCTGGCCAGGAAGTACATCTACTCGCTGTACTGGT 491

     225 erThrLeuIleLeuThrThrValGlyAspThrProProProAlaArgGlu 241
         ||||||||   |||||||||:::|||:::|||||||   |||::::::::
     492 CCACGCTGACCCTCACCACCATCGGAGAGACGCCTGCACCCGTCAAGGAT 541

     242 GluGluTyrLeuPheMetValGlyAspPheLeuLeuAlaValMetGlyPh 258
         ||||||||||:::::::::|||||||||   :::|||:::   ||
     542 ATCGAATACCTTTTCGTTATTTCGGACTTCCTCACCGGCGTGCTGATCTT 591

     258 eAlaThrIleMetGlySerMetSerSerValIleTyrAsnMetAsnThrA 275
         ||||:::|||:::|||:::::::::::::::|||   |||||||||:::|
     592 TGCTAGTATCGTCGGAAACGTCGGTGCCATGATCTCAAACATGAATGCTT 641

     275 laAspAlaAlaPheTyrProAspHisAlaLeuValLysLysTyrMetLys 291
         ::   |||   |||                  :::|||:::||||||:::
     642 CTCGGGCCGAGTTTCAGGCAAAAATCGACTCCATAAAGCAGTACATGCAG 691

     292 LeuGlnHisValAsnArgLysLeuGluArgArgValIleAspTrpTyrGl 308
         ::::::   |||:::::::::||||||   |||||||||   |||:::::
     692 TTCCGAAAGGTCTCCAAAGAACTGGAGGCCAGGGTCATCAAGTGGTTCGA 741

     308 nHisLeuGlnIleAsnLysLysMetThrAsnGluValAlaIleLeuGlnH 325
         ::::|||      :::||||||      :::|||      :::|||::::
     742 CTACCTGTGGACGGAGAAGAAGACCTGCGATGAGAAGGAAGTCCTGAAGA 791

     325 isLeuProGluArgLeuArgAlaGlu 333
         ::||||||:::::::|||:::||||||
     792 ACCTCCCGGACAAGCTCAAAGCCGAG 817
```

```
seq_name: gb_gss:AZ503692

seq_documentation_block:
LOCUS       AZ503692                621 bp    DNA     linear   GSS 05-OCT-2000
DEFINITION  1M0343L05R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0343L05 R, DNA sequence.
ACCESSION   AZ503692
VERSION     AZ503692.1  GI:10685008
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 621)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0343  row: L  column: 05
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 621.
FEATURES             Location/Qualifiers
     source          1. .621
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0343L05"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      143 a    166 c    123 g    189 t
ORIGIN


alignment_scores:
            Quality:  639.50              Length:    206
              Ratio:  3.675                Gaps:      2
 Percent Similarity:  84.466   Percent Identity: 59.709

alignment_block:
```

```
US-09-927-267-1 x AZ503692/rev   ..

Align seg 1/1  to reverse of: AZ503692  from: 1  to: 621

    333 GluValAlaValSerValHisLeuSerThrLeuSerArgValGlnIlePh 349
        |||:::|||:::::::|||||||||:::||||||:::::::|||:::|||||
    621 GAAATTGCCATCAACGTTCATCTAGACACGTTAAAAAAGGTTCGCATCTT 572

    349 eGlnAsnCysGluAlaSerLeuLeuGluGluLeuValLeuLysLeuGlnP 366
        |   :::|||||||||:::||||||   ||||||||||||||||||||||
    571 TGCTGACTGTGAAGCTGGTCTGTTGGTGGAGTTGGTGTTGAAACTACAAC 522

    366 roGlnThrTyrSerProGlyGluTyrValCysArgLysGlyAspIleGly 382
        |||||:::|||||||||||||:::|||:::|||:::|||||||||||||||
    521 CCCAGGTGTATAGTCCCGGAGATTACATATGCAAGAAAGGGGACATCGGG 472

    383 GlnGluMetTyrIleIleArgGluGlyGlnLeuAlaValValAlaAspAs 399
        :::|||||||||||||||:::||||||:::||||||||||||||||||||
    471 CGGGAGATGTACATCATCAAGGAAGGCAAACTTGCTGTGGTGGCAGATGA 422

    399 pGlyIleThrGlnTyrAlaValLeuGlyAlaGlyLeuTyrPheGlyGluI 416
        ||||||||||||:::::|||||:::   |||   ||||||||||||
    421 CGGAATTACACAGTTTGTGGTGTTGAGTGACGGCAGTTACTTTGGTGAGA 372

    416 leSerIleIleAsnIleLysGlyAsnMetSerGlyAsnArgArgThrAla 432
        ||||||||:::|||||||||||||:::   :::||||||||||||||||||
    371 TCAGCATTCTTAATATCAAAGGCAGCAAGGCTGGCAACCGAAGAACAGCC 322

    433 AsnIleLysSerLeuGlyTyrSerAspLeuPheCysLeuSerLysGluAs 449
        ||||||||||||:::|||||||||||||||||||||||||||||||:::||
    321 AACATTAAGAGCATCGGCTACTCGGACCTGTTCTGCCTCTCCAAAGATGA 272

    449 pLeuArgGluValLeuSerGluTyrProGlnAlaGlnThrIleMetGluG 466
        ||||   |||:::|||:::|||||||||:::|||:::|||:::::::|||
    271 CCTCATGGAAGCTCTTACAGAGTACCCAGATGCCAAAACTATGCTGGAAG 222

    466 luLysGlyArgGluIleLeuLeuLysMetAsnLysLeuAspValAsn... 481
        ||||||||||:::|||||||:::|||   :::   |||||||:::|||
    221 AGAAAGGGAGGCAGATCTTAATGAAAGATGGTCTGCTGGATATAAACATT 172

    482 ............AlaGluAlaAlaGluIleAlaLeuGlnGluAlaThrGl 494
                    ::::::      ::::::                 ||
    171 GCAAATATGGGCAGTGACCCTAAAGACCTG.................GA 140

    494 uSerArgLeuArgGlyLeuAspGlnGlnLeuAspAspLeuGlnThrLysP 511
        |::::::::::     ::::::   :::::::|||   |||||||||:::|
    139 AGAGAAGGTCACTCGAATGGAGGGGTCAGTGGACCTCCTGCAAACGCGAT 90

    511 heAlaArgLeuLeuAlaGluLeuGluSerSerAlaLeuLysIleAlaTyr 527
        ||||||||:::||||||||||   ||||||          |||:::
     89 TTGCCCGTATCTTGGCTGAATATGAGTCGATGCAGCAGAAACTCAAGCAA 40

    528 ArgIleGluArgLeuGlu 533
        |||:::   :::::::|||
     39 AGATTAACCAAGGTCGAG 22

seq_name: gb_est2:BF190904

seq_documentation_block:
LOCUS       BF190904                559 bp    mRNA    linear   EST 02-NOV-2000
DEFINITION  237569 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BF190904
VERSION     BF190904.1  GI:11074273
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 559)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 65  row: J  column: 10
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .559
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 2PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from testis, ovary,
                     endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT      184 a     94 c    136 g    145 t
ORIGIN


alignment_scores:
            Quality:  634.00             Length:    175
              Ratio:  3.938                Gaps:      0
 Percent Similarity:  92.000   Percent Identity: 66.286

alignment_block:
 US-09-927-267-1 x BF190904   ..

 Align seg 1/1  to: BF190904  from: 1  to: 559

    307 TyrGlnHisLeuGlnIleAsnLysLysMetThrAsnGluValAlaIleLe 323
        ::::::::::|||      |||||||||   ::::::|||      :::||
      1 TTTGACTATCTGTGGACCAACAAAAAAACAGTAGATGAGAAAGAAGTCTT 50

    323 uGlnHisLeuProGluArgLeuArgAlaGluValAlaValSerValHisL 340
        |:::::|||||:::::||||||||||||:::|||:::::|||||||
     51 GAAGTATCTACCTGATAAACTAAGAGCAGAAATTGCCATTAACGTTCACT 100

    340 euSerThrLeuSerArgValGlnIlePheGlnAsnCysGluAlaSerLeu 356
        ||:::||||||:::::::|||:::||||||   :::|||||||||:::|||
    101 TAGACACATTAAAAAAGGTGCGTATTTTTGCAGACTGTGAGGCTGGTCTC 150

    357 LeuGluGluLeuValLeuLysLeuGlnProGlnThrTyrSerProGlyGl 373
        |||   |||||||||||||||||||||||||:::|||||||||||||::
    151 TTGGTGGAGTTGGTCTTGAAATTACAGCCCCAGGTCTACAGTCCTGGAGA 200

    373 uTyrValCysArgLysGlyAspIleGlyGlnGluMetTyrIleIleArgG 390
        :|||:::|||:::|||||||||||||||:::|||||||||||||||:::|
    201 TTACATTTGCAAGAAAGGGGATATTGGACGAGAGATGTACATTATCAAAG 250

    390 luGlyGlnLeuAlaValValAlaAspAspGlyIleThrGlnTyrAlaVal 406
        |||||:::|||||||||||||||||||||||||||||||||:::::|||
    251 AAGGCAAACTTGCTGTGGTGGCAGATGATGGAATCACCCAGTTTGTGGTA 300

    407 LeuGlyAlaGlyLeuTyrPheGlyGluIleSerIleIleAsnIleLysGl 423
        |||:::   |||   |||||||||||||||||||:::||||||||||
    301 TTGAGTGATGGCAGCTACTTTGGTGAGATCAGCATCCTTAACATTAAAGG 350

    423 yAsnMetSerGlyAsnArgArgThrAlaAsnIleLysSerLeuGlyTyrS 440
        |:::   :::|||||||||||||||||||||||||||||:::|||||||
    351 CAGTAAAGCTGGCAATCGAAGAACAGCCAATATTAAAAGTATTGGCTATT 400

    440 erAspLeuPheCysLeuSerLysGluAspLeuArgGluValLeuSerGlu 456
```

```
     |||||||||||||||||||||||:::||||||   |||:::|||:::|||
401 CAGATCTGTTCTGTCTCTCGAAAGATGACCTCATGGAAGCTCTAACCGAG 450

457 TyrProGlnAlaGlnThrIleMetGluGluLysGlyArgGluIleLeuLe 473
    ||||||:::|||:::   ::::::||||||||||||||:::::||||||::
451 TACCCAGATGCTAAATGTATGCTAGAAGAGAAAGGGAAGCAAATCTTGAT 500

473 uLysMetAsnLysLeuAspValAsn 481
    :|||   :::   ||||||:::|||
501 GAAAGATGGTCTGCTGGATATCAAT 525

seq_name: gb_gss:AQ537904

seq_documentation_block:
LOCUS       AQ537904                627 bp    DNA     linear   GSS 18-MAY-1999
DEFINITION  RPCI-11-326D9.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-326D9,
            DNA sequence.
ACCESSION   AQ537904
VERSION     AQ537904.1  GI:4849594
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 627)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI-11-326D9.TV
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .627
                     /organism="Homo sapiens"
                     /db_xref="GDB:7624880"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-326D9"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      151 a    164 c    157 g    155 t
ORIGIN


alignment_scores:
            Quality:  634.00             Length:    206
              Ratio:  3.665                Gaps:      0
 Percent Similarity:  83.981   Percent Identity: 54.854

alignment_block:
 US-09-927-267-1 x AQ537904   ..

 Align seg 1/1  to: AQ537904  from: 1  to: 627

    152 GluAlaPheAspArgThrGluThrArgThrAlaTyrProAsnAlaPheAr 168
        |||   |||||||||||||||||||||||   |||||||||   |||||

      8 GAATTCTTTGACCGCACAGAGACAAGGACCAACTACCCCAATATGTTCAG 57

    168 gIleAlaLysLeuMetLeuTyrIlePheValValIleHisTrpAsnSerC 185
        ||||:::::::|||:::|||||||||:::::::::|||||||||||:::|
     58 GATTGGTAACTTGGTCTTGTACATTCTCATCATCATCCACTGGAATGCCT 107

    185 ysLeuTyrPheAlaLeuSerArgTyrLeuGlyPheGlyArgAspAlaTrp 201
        ||:::||||||||||:::|||:::::::::|||||||||   |||:::|||
    108 GCATCTACTTTGCCATTTCCAAGTTCATTGGTTTTGGGACAGACTCCTGG 157

    202 ValTyrProAspProAlaGlnProGlyPheGluArgLeuArgArgGlnTy 218
        |||||||||:::   :::   |||         ||||||   |||:::||
    158 GTCTACCCAAACATCTCAATCCCAGAGCATGGGCGCCTCTCCAGGAAGTA 207

    218 rLeuTyrSerPheTyrPheSerThrLeuIleLeuThrThrValGlyAspT 235
        |:::||||||:::|||:::|||||||||   |||||||||:::|||:::|
    208 CATTTACAGTCTCTACTGGTCCACCTTGACCCTTACCACCATTGGTGAGA 257

    235 hrProProProAlaArgGluGluGluTyrLeuPheMetValGlyAspPhe 251
        ||||||||||:::::::::|||||||||||||||||||:::   ||||||
    258 CCCCACCCCCCGTGAAAGATGAGGAGTATCTCTTTGTGGTCGTAGACTTC 307

    252 LeuLeuAlaValMetGlyPheAlaThrIleMetGlySerMetSerServa 268
        |||:::::::|||:::   ||||||||||||:::|||:::::::::|||::
    308 TTGGTGGGTGTTCTGATTTTTGCCACCATTGTGGGCAATGTGGGCTCCAT 357

    268 lIleTyrAsnMetAsnThrAlaAspAlaAlaPheTyrProAspHisAlaL 285
        :|||   |||||||||:::::   |||   |||
    358 GATCTCGAATATGAATGCCTCACGGGCAGAGTTCCAGGCCAAGATTGATT 407

    285 euValLysLysTyrMetLysLeuGlnHisValAsnArgLysLeuGluArg 301
          :::|||:::||||||:::::::::   |||::::::   ||||||
    408 CCATCAAGCAGTACATGCAGTTCCGCAAGGTCACCAAGGACTTGGAGACG 457

    302 ArgValIleAspTrpTyrGlnHisLeuGlnIleAsnLysLysMetThrAs 318
        |||||||||   |||:::::::::|||      |||||||||   :::::
    458 CGGGTTATCCGGTGGTTTGACTACCTGTGGGCCAACAAGAAGACGGTGGA 507

    318 nGluValAlaIleLeuGlnHisLeuProGluArgLeuArgAlaGluValA 335
        :|||      :::|||:::   ||||||:::::::|||:::|||||||:::|
    508 TGAGAAGGAGGTGCTCAAGAGCCTCCCAGACAAGCTGAAGGCTGAGATCG 557

    335 laValSerValHisLeuSerThrLeuSerArgValGlnIlePheGlnAsn 351
        ||:::::|||||||||:::||||||:::::|||:::|||||||||:::
    558 CCATCAACGTGCACCTGGACACGCTGAAGAAGGTTCGCATCTTCCAGGAC 607

    352 CysGluAlaSerLeuLeu 357
        |||||||||   ||||||
    608 TGTGAAGCAAGGCTGCTG 625

seq_name: gb_est1:AW640105

seq_documentation_block:
LOCUS       AW640105                558 bp    mRNA    linear   EST 26-APR-2001
DEFINITION  bl91d05.w1 Blackshear/Soares normalized Xenopus egg library Xenopus
            laevis cDNA clone PBX0091D05 5', mRNA sequence.
ACCESSION   AW640105
VERSION     AW640105.1  GI:7397317
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 558)
  AUTHORS   Blackshear,P.J., Lai,W.S., Thorn,J.M., Kennington,E.A., Staffa,N.G.
            Jr., Moore,D.T., Bouffard,G.G., Beckstrom-Sternberg,S.M., Touchman
            ,J.W., Bonaldo,M.F. and Soares,M.B.
  TITLE     The NIEHS Xenopus maternal EST project: interim analysis of the
            first 13,879 ESTs from unfertilized eggs
  JOURNAL   Gene 267 (1), 71-87 (2001)
  MEDLINE   21211403
COMMENT     Contact: Perry J. Blackshear
```

```
            Office of Clinical Research and Laboratory of Signal Transduction
            National Institute of Environmental Health Sciences
            A2-05 NIEHS, 101 Alexander Drive, Research Triangle Park, NC 27709,
            USA
            Tel: 919 541-4899
            Fax: 919 541-4571
            Email: black009@niehs.nih.gov
            Clone is available through Research Genetics, Inc., 2130 Memorial
            Parkway, Huntsville, AL 35901
             phone 800-533-4363 ext.cdna, fax 256-536-9016 att:cdna, email
            cdna@resgen.com
             DNA Sequencing and analyses performed by National Institutes of
            Health Intramural Sequencing Center (NISC).
            PCR PRimers
            FORWARD: TGTAAAACGACGGCCAGT
            BACKWARD: CAGGAAACAGCTATGACC
            Plate: 0091  row: D  column: 05
            Seq primer: T7 primer.
FEATURES              Location/Qualifiers
     source           1. .558
                      /organism="Xenopus laevis"
                      /db_xref="taxon:8355"
                      /clone="PBX0091D05"
                      /clone_lib="Blackshear/Soares normalized Xenopus egg
                      library"
                      /sex="female"
                      /tissue_type="unfertilized egg"
                      /cell_type="unfertilized egg"
                      /dev_stage="unfertilized egg"
                      /lab_host="DH10B"
                      /note="Vector: pT7T3-Pac; Site_1: EcoRI; Site_2: NotI;
                      PolyA-selected mRNA was prepared from unfertilized Xenopus
                      laevis eggs.  The library was constructed in the vector
                      pT7T3-Pac as described in Bonaldo, M.F., Lennon, G. and
                      Soares, M.B. 'Normalization and subtraction:  two
                      approaches to facilitate gene discovery',  Genome Research
                      6:791-806, 1996.  The first strand synthesis used a
                      NotI-dT18 primer; double stranded cDNAs were ligated to
                      EcoRI adapters, digested with NotI, and directionally
                      cloned into the NotI and EcoRI-digested pT7T3-Pac vector.
                      The library contained approximately 7.2 X 10^5
                      recombinants, with average insert sizes of 1-1.5 kb."
BASE COUNT      191 a     97 c    136 g    133 t      1 others
ORIGIN


alignment_scores:
           Quality:  623.50            Length:    185
             Ratio:  3.825               Gaps:      1
 Percent Similarity:  88.108   Percent Identity: 63.784

alignment_block:
 US-09-927-267-1 x AW640105   ..

 Align seg 1/1  to: AW640105  from: 1  to: 558

     301 ArgArgValIleAspTrpTyrGlnHisLeuGlnIleAsnLysLysMetTh 317
         |||:::   :::|||         :::|||      |||||||||
           2 CGGCACGAGGTTGAC.........TACCTCTGGACAAACAAAAAAACCAT 42

     317 rAsnGluValAlaIleLeuGlnHisLeuProGluArgLeuArgAlaGluV 334
          :::|||      :::|||:::   ||||||:::::|||:::||||||:
      43 GGATGAGAAGGAGGTATTAAAGAGCTTGCCAGACAAATTAAAGGCTGAAA 92

     334 alAlaValSerValHisLeuSerThrLeuSerArgValGlnIlePheGln 350
         ::|||:::::|||||||||:::||||||:::::::|||:::||||||||
      93 TTGCAATTAATGTACATTTAGACACTTTAAAAAAAGTTAGGATCTTTCAA 142

     351 AsnCysGluAlaSerLeuLeuGluGluLeuValLeuLysLeuGlnProGl 367
         :::|||||||||:::||||||   |||||||||||||||||:::||||||
     143 GACTGCGAGGCTGGATTGCTTATAGAACTGGTACTGAAACTAAGGCCCCA 192

     367 nThrTyrSerProGlyGluTyrValCysArgLysGlyAspIleGlyGlnG 384
         |:::::|||||||||:::|||:::|||||||||||||||||||||:::|
     193 AGTCTTTAGCCCTGGTGACTACATTTGCCGAAAAGGAGATATTGGAAGGG 242

     384 luMetTyrIleIleArgGluGlyGlnLeuAlaValValAlaAspAspGly 400
         |||||||||||||:::|||||||:::|||||||||||||||||||||||
     243 AAATGTACATCATCAAAGAGGGGAAACTAGCTGTAGTAGCTGATGATGGC 292

     401 IleThrGlnTyrAlaValLeuGlyAlaGlyLeuTyrPheGlyGluIleSe 417
         ||||||||:::::|||||:::   |||   ||||||||||||||||||
     293 ATCACTCAGTTTGTGGTCCTCAGTGATGGCAGTTATTTTGGCGAAATTAG 342

     417 rIleIleAsnIleLysGlyAsnMetSerGlyAsnArgArgThrAlaAsnI 434
         ||||:::|||||||||||||:::   ||||||||||||||||||||||||
     343 CATCTTAAACATCAAAGGAAGCAAGTCAGGCAACCGGAGGACAGCCAACA 392

     434 leLysSerLeuGlyTyrSerAspLeuPheCysLeuSerLysGluAspLeu 450
         ||:::|||:::|||||||||||||||||||||||||||||||:::||||||
     393 TCAGAAGTATTGGCTATTCCGACCTGTTTTGTTTATCAAAGGATGATTTG 442

     451 ArgGluValLeuSerGluTyrProGlnAlaGlnThrIleMetGluGluLy 467
           |||:::|||:::|||||||||:::|||:::|||::::::|||||||
     443 ATGGAGGCACTTACTGAATATCCTGATGCTAAAACAGTCTTAGAAGAAAA 492

     467 sGlyArgGluIleLeuLeuLysMetAsnLysLeuAspValAsnAlaGluA 484
         |||||||:::|||   :::|||   :::   :::|||   :::|||
     493 AGGGCGACAGATCCNCATGAAAGATGGTTTAATAGATGAAGAAGCAGCCA 542

     484 laAla 485
           |||
     543 AAGCA 547

seq_name: gb_est2:BF190899

seq_documentation_block:
LOCUS       BF190899                560 bp    mRNA    linear   EST 02-NOV-2000
DEFINITION  237563 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BF190899
VERSION     BF190899.1  GI:11074268
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 560)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 65  row: I  column: 10
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES              Location/Qualifiers
     source           1. .560
                      /organism="Sus scrofa"
                      /db_xref="taxon:9823"
                      /clone_lib="MARC 2PIG"
                      /tissue_type="pooled"
                      /lab_host="DH10B"
                      /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                      Library made from pooled tissue from testis, ovary,
```

endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT 183 a 95 c 136 g 146 t
ORIGIN

alignment_scores:
Quality: 621.00 Length: 175
Ratio: 3.881 Gaps: 0
Percent Similarity: 91.429 Percent Identity: 65.143

alignment_block:
US-09-927-267-1 x BF190899 ..

Align seg 1/1 to: BF190899 from: 1 to: 560

```
307 TyrGlnHisLeuGlnIleAsnLysLysMetThrAsnGluValAlaIleLe 323
    ::::::::::|||      |||||||||   :::::::|||      :::|| 
  1 TTTGACTATCTGTGGACCAACAAAAAAACAGTAGATGAGAAAGAAGTCTT 50

323 uGlnHisLeuProGluArgLeuArgAlaGluValAlaValSerValHisL 340
    |::::::|||||:::::|||||||||||||:::|||:::::::|||||||
 51 GAAGTATCTACCTGATAAACTAAGAGCAGAAATTGCCATTAACGTTCACT 100

340 euSerThrLeuSerArgValGlnIlePheGlnAsnCysGluAlaSerLeu 356
    ||:::|||||:::::|||:::|||||   :::|||||||||:::|||
101 TAGACACATTAAAAAAGGTGCGTATTTTTGCAGACTGTGAGGCTGGTCTC 150

357 LeuGluGluLeuValLeuLysLeuGlnProGlnThrTyrSerProGlyGl 373
    |||   ||||||||||||||||||||||||||:::||||||||||||::
151 TTGGTGGAGTTGGTCTTGAAATTACAGCCCCAGGTCTACAGTCCTGGAGA 200

373 uTyrValCysArgLysGlyAspIleGlyGlnGluMetTyrIleIleArgG 390
    :|||:::|||:::|||||||||||||||||:::|||||||||||||||:::|
201 TTACATTTGCAAGAAAGGGGATATTGGACGAGAGATGTACATTATCAAAG 250

390 luGlyGlnLeuAlaValValAlaAspAspGlyIleThrGlnTyrAlaVal 406
    |||||:::|||||||||||||||||||||||||||||:::::::::|||
251 AAGGCAAACTTGCTGTGGTGGCAGATGATGGAATCACCCACTTTGTGGTA 300

407 LeuGlyAlaGlyLeuTyrPheGlyGluIleSerIleIleAsnIleLysGl 423
    |||:::   |||   |||||||||||||||||||:::||||||||||||
301 TTGAGTGATGGCAGCTACTTTGGTGAGATCAGCATCCTTAACATTAAAGG 350

423 yAsnMetSerGlyAsnArgArgThrAlaAsnIleLysSerLeuGlyTyrS 440
    |:::   :::|||   |||||||||||||||||||||||:::|||||||
351 CAGTAAAGCTGGCTATCGAAGAACAGCCAATATTAAAAGTATTGGCTATT 400

440 erAspLeuPheCysLeuSerLysGluAspLeuArgGluValLeuSerGlu 456
    |||||||||||||||||||||:::||||||   |||:::|||:::|||
401 CAGATCTGTTCTGTCTCTCGAAAGATGACCTCATGGAAGCTCTAACCGAG 450

457 TyrProGlnAlaGlnThrIleMetGluGluLysGlyArgGluIleLeuLe 473
    ||||||:::|||:::   ::::::||||||||||||:::::::||||||::
451 TACCCAGATGCTAAATGTATGCTAGAAGAGAAAGGGAAGCAAATCTTGAT 500

473 uLysMetAsnLysLeuAspValAsn 481
    :|||   :::   ||||||:::|||
501 GAAAGATGGTCTGCTGGATATCAAT 525
```

seq_name: gb_est1:BB625423

seq_documentation_block:
LOCUS BB625423 646 bp mRNA linear EST 26-OCT-2001
DEFINITION BB625423 RIKEN full-length enriched, adult male epididymis Mus musculus cDNA clone 9230103C20 5', mRNA sequence.
ACCESSION BB625423
VERSION BB625423.1 GI:16463693
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 646)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.
FEATURES Location/Qualifiers
source 1. .646
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="9230103C20"
/clone_lib="RIKEN full-length enriched, adult male epididymis"
/sex="male"
/tissue_type="epididymis"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"

```
BASE COUNT      146 a    221 c    170 g    109 t
ORIGIN


alignment_scores:
            Quality:  589.00             Length:    170
              Ratio:  4.119                Gaps:      2
 Percent Similarity:  84.118   Percent Identity: 70.588

alignment_block:
 US-09-927-267-1 x BB625423   ..

 Align seg 1/1  to: BB625423  from: 1  to: 646

       1 MetSerGlnAspThrLysValLysThrThrGluSerSerProProAlaPr 17
         ||||||||||:|||:|||||||||||||||||||||||:||:|||||||||||
     127 ATGAGCCAGGACAGCAAAGTGAAGACAACAGAGTCCACGCCCCCAGCCCC 176

      17 oSerLysAlaArgLysLeuLeuProValLeuAspProSerGlyAspTyrT 34
         |:::|||||||||||||   |||||||||||||||||||||||||||||||
     177 AACCAAAGCCAGGAAGCGGCTGCCAGTCCTAGACCCATCTGGGGATTATT 226

      34 yrTyrTrpTrpLeuAsnThrMetValPheProValMetTyrAsnLeuIle 50
         ||||||||   |||||||||   :::||||||:::|||||||||||||||
     227 ACTACTGGGGGCTGAACACGAGGGCCTTCCCAATCATGTATAATCTCATC 276

      51 IleLeuValCysArgAlaCysPheProAspLeuGlnHisGlyTyrLeuVa 67
         |||   :::|||||||||||||||||||||||||||||      ||||||||
     277 ATCGCTGCATGCAGGGCCTGCTTTCCTGACTTGCAACCCAGATACCTGGT 326

      67 lAlaTrpLeuValLeuAspTyrThrSerAspLeuLeuTyrLeuLeuAspM 84
         ||||||||||   ||||||||||||   |||||||||||||||||||||:
     327 GGCCTGGCTTGGGCTGGACTACACGAGGGACCTGCTGTACCTACTAGATG 376

      84 etValValArgPheHisThrGlyPheLeuGluGlnGly.IleLeuValVa 100
         ::   |||||||||||||||||||||||||||||||| :::||||||
     377 TTGGGGTACGCTTCCACACAGGATTCCTAGAACAGGGCACTCCTGGTGGG 426

     100 lAspLysGlyArgIleSerSerArgTyrValArgThrTrpSerPhePheL 117
          ||||||:::   |||:::   ||||||   |||      |||:::::::
     427 GGACAAAAGCACGATCGCCCGGCGCTATGGTCGCCCCCGGAGCCTCCTGG 476

     117 euAspLeuAlaSerLeuMetProThrAspValValTyrValArgLeuGly 133
         ::|||   ||||||||||:::|||   |||:::::|||||||:::||||||
     477 TGGACCCGGCTTCCCTGGTCCCCCCAGATGCAGCCTACGTGCAGTTAGGC 526

     134 ProHisThrProThrLeuArgLeuAsnArgPheLeuArg.AlaProArgL 150
         |||||||||   |||:::|||||||||   |||   ||| :::|||||||
     527 CCGCACACCCGTACCGTCCGGCTAAACCCGTTCCCCCGCAGGCCCTCGCC 576

     150 euPheGluAlaPheAspArgThrGluThrArgThrAlaTyrProAsnAla 166
         ||      :::      ||||||   ||||||||||||:::|||:::|||
     577 TCTGTGCGGGTCCCGCCCGGACCGGGACCCGAACCGCTCACCCACATGCC 626

     167 PheArg 168
         :::|||
     627 CTCCGC 632

seq_name: gb_est2:BG342539

seq_documentation_block:
LOCUS       BG342539              1019 bp    mRNA    linear   EST 27-FEB-2001
DEFINITION  602374710F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4482002 5',
            mRNA sequence.
ACCESSION   BG342539
VERSION     BG342539.1  GI:13148977
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1019)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10318  row: d  column: 03
            High quality sequence stop: 642.
FEATURES             Location/Qualifiers
     source          1. .1019
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4482002"
                     /clone_lib="NIH_MGC_94"
                     /tissue_type="retina"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 3.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      344 a    215 c    233 g    227 t
ORIGIN


alignment_scores:
            Quality:  577.50             Length:    348
              Ratio:  2.661                Gaps:      9
 Percent Similarity:  62.356   Percent Identity: 36.782

alignment_block:
 US-09-927-267-1 x BG342539   ..

 Align seg 1/1  to: BG342539  from: 1  to: 1019

       6 LysValLysThrThrGluSerSerProProAlaProSerLysAlaArgLy 22
         |||   |||      |||:::::::            ::::::   :::||
      27 AAGAAAAAGAAGGAGGAGAAAACCAAAGAGAAGAAAGAAGAGGAGAAGAA 76

      22 sLeuLeuProValLeuAspProSerGlyAspTyrTyrTyrTrpTrpLeuA 39
         |   :::   |||:::|||||||||||||:::   ||||||   ||||||
      77 AGAAGTCGTGGTTATTGATCCTTCAGGAAACACATACTACAACTGGCTGT 126

      39 snThrMetValPheProValMetTyrAsnLeuIleIleLeuValCysArg 55
           :::::::::|||||||||||||||      ::::::::::|||
     127 TTTGTATCACTTTGCCTGTGATGTACAACTGGACGATGATTATTGCAAGA 176

      56 AlaCysPheProAspLeuGlnHisGlyTyrLeuValAlaTrpLeuValLe 72
         |||||||||   :::||||||      ||||||      ||||||:::
     177 GCATGTTTTGATGAACTCCAGTCTGATTACCTAGAATACTGGCTCATTTA 226

      72 uAspTyrThrSerAspLeuLeuTyrLeuLeuAspMetValValArgPheH 89
          ||||||:::||||||:::::|||||||   ||||||   ||||||   :
     227 TGATTACGTATCAGATGTAGTCTATCTTGCCGATATGTTTGTACGAACAA 276

      89 isThrGlyPheLeuGluGlnGlyIleLeuValValAspLysGlyArgIle 105
         ::|||||:::|||||||||||||:::||||||   |||:::   ::::::
     277 GGACAGGTTACCTGGAACAAGGGCTGCTAGTGAAAGACAGAATGAAACTC 326

     106 SerSerArgTyrValArgThrTrpSerPhePheLeuAspLeuAlaSerLe 122
           ::::::|||      :::   :::|||   ||||||:::   |||::
     327 ATAGAGAAATACAAAGCAAACCTGCAGTTTAAACTTGATGTTCTGTCAGT 376

     122 uMetProThrAspValValTyrValArgLeuGlyProHisThrProThrL 139
         ::::|||||||||:::::::|||:::::   |||   :::   |||   :
     377 GATACCGACCGATCTGCTGTACATCAAGTATGGGTGGAACTATCCAGAAA 426

     139 euArgLeuAsnArgPheLeuArgAlaProArgLeuPheGluAlaPheAsp 155
```

```
     ::|||||||||||||:::||||||      |||:::||||||   |||:::
427 TCAGGTTAAACCGGCTGTTAAGGATCTCTCGAATGTTTGAGTTCTTCCAG 476

156 ArgThrGluThrArgThrAlaTyrProAsnAlaPheArgIleAlaLysLe 172
    ||||||||||||||||   |||||||||   |||||||||::::::|||
477 AGGACGGAAACAAGGACCAACTACCCAAACATCTTTAGGATTTCGAACCT 526

172 uMetLeuTyrIlePheValValIleHisTrpAsnSerCysLeuTyrPheA 189
    |::::::||||||   ::::::|||||||||||||:::|||:::|||:::
527 TGTGATGTACATCGTCATCATCATCCACTGGAACGCTTGTGTGTATTATT 576

189 laLeuSer.ArgTyrLeuGlyPheGlyArgAspAlaTrpValTyrProAs 205
    :::::||| :::   :::|||||||||:::|||:::|||||||||||||||
577 CCATCTCAAAAAGCTATTGGATTCGGGAATGACACATGGGTCTACCCTGA 626

205 pProAlaGlnProGlyPheGluArgLeuArgArgGlnTyrLeuTyrSerP 222
    |      :::|||            :::|||||||      |||   ::::
627 TGTTAATGATCCTGA.....ATCCGGCGGTATAGCTTAGAAGAAT 666

222 heTyrPheSerThrLeuIleLeuThrThrValGlyAspThrProPro... 237
    ::   ::::::   ||||||   :::   :::|||          |||
667 ATTGTCTACAGCCTTTAATAGGGTCTAAATTAGGCTTTGGGAAACCCATG 716

238 ..............ProAlaArgGluGluGluTyrLeuPheMetValGl 249
                  |||      ::::::   |||::::::      ||
717 GGTGGAACTCAACTTCCGTGGCTGGATTCTGGGTACATCTATGCGGATGG 766

249 yAspPheLeuLeuAlaValMetGlyPheAlaThrIleMetGlySerMetS 266
    ||||:::|||:::::::::
767 AGACATCCTCATCGGGACTTCA..........................  788

266 erSerValIleTyrAsnMetAsnThrAlaAspAlaAlaPheTyrProAsp 282
                                                |||   :::
789 ....................................TATTCGCAA 797

283 HisAlaLeuValLysLysTyr.MetLysLeuGlnHisValAsnArgLysL 299
              :::   |||::: |||      :::|||:::   |||:::
798 CCAAGACGGATCCATAAGTTCCATGACTCCAAAACACATGCCCCGGCGCA 847

299 eu...GluArgArgValIleAspTrpTyrGlnHisLeuGlnIleAsnLys 314
        :::|||||           |||         :::   |||:::
848 ATTGAGACAGAAGG...........TACGCATCAACACGACAGAACCGC 885

315 LysMetThrAsnGluValAlaIle..................... 322
    |||      |||::::::|||
886 AAG......AACCAGATAGCGACACGAGATCAACGGCTACAACCCCAGAA 929

323 .........LeuGlnHisLeuProGluArgLeuArgAla 332
            :::|||||:::   :::|||:::::|||
930 CACCACACAGGTACAACATGTGAAACAGAGAATACAAGCG 969
```

seq_name: gb_est2:BF937274

seq_documentation_block:

```
LOCUS       BF937274                567 bp    mRNA    linear   EST 13-FEB-2002
DEFINITION  fm62h03.y1 Zebrafish adult retina cDNA Danio rerio cDNA clone
            4200053 5' similar to SW:CNG1_BOVIN Q00194 CGMP-GATED CATION
            CHANNEL PROTEIN ;, mRNA sequence.
ACCESSION   BF937274
VERSION     BF937274.1  GI:12354594
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 567)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            Library constructed by: Chandra Tucker and Gregory Niemi DNA
            Sequencing by: Washington University Genome Sequencing Center Clone
            distribution: RessourcenZentrumPrimarDatenbank, Berlin, Germany
            (web address: www.rzpd.de)
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 420.
FEATURES             Location/Qualifiers
     source          1. .567
                     /organism="Danio rerio"
                     /strain="wild-type"
                     /db_xref="taxon:7955"
                     /clone="4200053"
                     /clone_lib="Zebrafish adult retina cDNA"
                     /sex="mixed"
                     /dev_stage="1-2 years"
                     /lab_host="E.Coli XL1-Blue MRF' (XL1-Blue MRF')"
                     /note="Vector: Lambda ZAP II (pBluescript SK-); Site_1:
                     EcoRI; Site_2: SalI; This Zebrafish library was
                     constructed by Dr. Susan E. Brockerhoff (email:
                     sbrocker@u.washington.edu) RZPD library number: 760"
BASE COUNT      156 a    133 c    143 g    134 t      1 others
ORIGIN
```

alignment_scores:

```
            Quality:  559.00            Length:    189
              Ratio:  3.494               Gaps:      0
 Percent Similarity:  84.656   Percent Identity: 54.497
```

alignment_block:

US-09-927-267-1 x BF937274 ..

Align seg 1/1 to: BF937274 from: 1 to: 567

```
203 TyrProAspProAlaGlnProGlyPheGluArgLeuArgArgGlnTyrLe 219
    ||||||||||||:::::|||   |||   ||||||   |||:::|||
  1 TACCCGGACCCTTCAGATCCTGAGTTTGGCCGATTGGTACGTAAATATGC  50

219 uTyrSerPheTyrPheSerThrLeuIleLeuThrThrValGlyAspThrP 236
     ||||||:::|||:::|||||||||   |||||||||:::|||:::||||
 51 CTACAGCATGTACTGGTCCACCCTGACGCTCACTACCATCGGCGAAACGC 100

236 roProProAlaArgGluGluGluTyrLeuPheMetValGlyAspPheLeu 252
    ||||||||:::::::::::::|||||:::|||:::|||   ||||||||
101 CGCCGCCCGTCGAAAACTCTGAGTACTTCTTCGTAGTCACCGACTTTCTA 150

253 LeuAlaValMetGlyPheAlaThrIleMetGlySerMetSerSerValIl 269
    ::::::|||:::   ||||||||||||:::|||:::::::::|||:::||
151 GTAGGCGTGTTGATTTTCGCAACAATCGTCGGTAACGTCGGCTCAATGAT 200

269 eTyrAsnMetAsnThrAlaAspAlaAlaPheTyrProAspHisAlaLeuV 286
    |   |||   |||:::|||   |||   |||                  :
201 CACAAATGCCAACGCAGCACGAGCCGATTTCCAAGCTCGCATCGATGCTA 250

286 alLysLysTyrMetLysLeuGlnHisValAsnArgLysLeuGluArgArg 302
    ::|||:::||||||:::::::::   |||::::::   ||||||:::|||
251 TTAAACAATACATGAGCTTTCGGAAAGTCACCAAGGATTTGGAGAAACGG 300

303 ValIleAspTrpTyrGlnHisLeuGlnIleAsnLysLysMetThrAsnGl 319
    ||||||   |||:::::::::|||      |||||||||   ::::::||
301 GTTATTAAATGGTTTGATTATTTGTGGACCAATAAAAAGGCGGTCGATGA 350

319 uValAlaIleLeuGlnHisLeuProGluArgLeuArgAlaGluValAlaV 336
    |      :::|||:::::::||||||:::::|||||||||||||:::|||:
351 GAGAGAAGTACTGAAGTATCTTCCCGACAAGCTGAGGGCGGAGATAGCCA 400
```

```
     336 alSerValHisLeuSerThrLeuSerArgValGlnIlePheGlnAsnCys 352
         :::::|||||||||:::||||||:::::::|||:::|||      :::|||
     401 TAAACGTCCATCTGGATACGCTAAAAAAAGTGCGTATCTNTGCGGACTGT 450

     353 GluAlaSerLeuLeuGluGluLeuValLeuLysLeuGlnProGlnThrTy 369
         ||||||:::||||||   ||||||||||||||||||||||||||||:::||
     451 GAAGCCGGGCTTTTGGTGGAGCTGGTGTTAAAGCTGCAGCCGCAGGTTTA 500

     369 rSerProGlyGluTyrValCysArgLysGlyAspIleGlyGlnGluMetT 386
         ||||||||||:::|||:::|||::::::|||||||||   :::|||||||
     501 CAGCCCTGGAGACTATATATGCAAAAACGGCGACATCCGGCGCGAGATGT 550

     386 yrIleIleArgGluGly 391
         ||||||||:::::::|||
     551 ACATCATTAAAGACGGT 567

seq_name: gb_est2:BE491188

seq_documentation_block:
LOCUS       BE491188                580 bp    mRNA    linear   EST 31-JUL-2000
DEFINITION  db43c03.x1 Blackshear/Soares normalized Xenopus egg library Xenopus
            laevis cDNA clone IMAGE:3301252 3' similar to gb:S42457 CGMP-GATED
            CATION CHANNEL PROTEIN (HUMAN); gb:U19717 Mus musculus cGMP-gated
            cation channel mRNA, clone mIMCD-K2 (MOUSE);, mRNA sequence.
ACCESSION   BE491188
VERSION     BE491188.1  GI:9610649
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 580)
  AUTHORS   Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.
            , Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
            ,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     WashU Xenopus EST project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Sandy Clifton, Ph.D.
            WashU Xenopus EST project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library constructed by Bento Soares and M. Fatima Bonaldo
            (University of Iowa). DNA Sequencing by: Washington University
            Genome Sequencing Center
             Clone distribution: Xenopus clones from this library are available
            through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
            Seq primer: -40UP from Gibco
            High quality sequence stop: 453.
FEATURES             Location/Qualifiers
     source          1. .580
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="IMAGE:3301252"
                     /clone_lib="Blackshear/Soares normalized Xenopus egg
                     library"
                     /sex="female"
                     /tissue_type="unfertilized egg"
                     /cell_type="unfertilized egg"
                     /dev_stage="unfertilized egg"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3-Pac; Site_1: EcoRI; Site_2: NotI;
                     PolyA-selected mRNA was prepared from unfertilized Xenopus
                     laevis eggs.  The library was constructed in the vector
                     pT7T3-Pac as described in Bonaldo, M.F., Lennon, G. and
                     Soares, M.B. 'Normalization and subtraction:  two
                     approaches to facilitate gene discovery',  Genome Research
                     6:791-806, 1996.  The first strand synthesis used a
                     NotI-dT18 primer; double stranded cDNAs were ligated to
                     EcoRI adapters, digested with NotI, and directionally
                     cloned into the NotI and EcoRI-digested pT7T3-Pac vector.
                     The library contained approximately 7.2 X 10^5
                     recombinants, with average insert sizes of 1-1.5 kb."
BASE COUNT      130 a    148 c     95 g    207 t
ORIGIN


alignment_scores:
            Quality:  521.50              Length:    206
              Ratio:  3.343                 Gaps:      2
 Percent Similarity:  75.728    Percent Identity: 51.942

alignment_block:
 US-09-927-267-1 x BE491188/rev   ..

 Align seg 1/1  to reverse of: BE491188  from: 1  to: 580

     367 GlnThrTyrSerProGlyGluTyrValCysArgLysGlyAspIleGlyGl 383
         |||:::::::|||   |||:::|||:::|||||||||||||||||||||::
     580 CAAGTCTTTAGCGCTGGTGACTACATTTGCCGAAAAGGAGATATTGGAAG 531

     383 nGluMetTyrIleIleArgGluGlyGlnLeuAlaValValAlaAspAspG 400
         :|||||||||||||||:::||||||:::|||||||||||||||||||||||
     530 GGAAATGTACATCATCAAAGAGGGGAAACTAGCTGTAGTAGCTGATGATG 481

     400 lyIleThrGlnTyrAlaValLeuGlyAlaGlyLeuTyrPheGlyGluIle 416
         ||||||||||:::::|||||:::   |||   ||||||||||||||||
     480 GCATCACTCAGTTTGTGGTCCTCAGTGATGGCAGTTATTTTGGCGAAATT 431

     417 SerIleIleAsnIleLysGlyAsnMetSerGlyAsnArgArgThrAlaAs 433
         ||||||:::|||||||||||||:::   ||||||||||||||||||||||
     430 AGCATCTTAAACATCAAAGGAAGCAAGTCAGGCAACCGGAGGACAGCCAA 381

     433 nIleLysSerLeuGlyTyrSerAspLeuPheCysLeuSerLysGluAspL 450
         ||||:::|||:::|||||||||||||||||||||||||||||||:::||||
     380 CATCAGAAGTATTGGCTATTCCGACCTGTTTTGTTTATCAAAGGATGATT 331

     450 euArgGluValLeuSerGluTyrProGlnAlaGlnThrIleMetGluGlu 466
         ||   |||:::|||:::|||||||||:::|||:::|||:::::::||||
     330 TGATGGAGGCACTTACTGAATATCCTGATGCTAAAACAGTCTTAGAAGAA 281

     467 LysGlyArgGluIleLeuLeuLysMetAsnLysLeuAspValAsnAlaGl 483
         |||||||||:::|||||||:::|||   :::   :::|||   :::|||
     280 AAAGGGCGACAGATCCTCATGAAAGATGGTTTAATAGATGAAGAAGCAGC 231

     483 uAlaAlaGluIleAlaLeuGlnGluAlaThrGluSerArgLeuArgGlyL 500
         |||      |||      :::      |||:::::::::::::   :
     230 CAAAGCAGGT...GCAGACCCGAAGGATATAGAGGAAAAGGTGGAGAAGA 184

     500 euAspGlnGlnLeuAspAspLeuGlnThrLysPheAlaArgLeuLeuAla 516
         ::         :::|||:::||||||||||:::|||||||||||||:::|||
     183 TGGGAATTGTAATGGATAGCTTACAGACTCGATTTGCCAGGCTAATGGCC 134

     517 GluLeuGluSerSerAlaLeuLysIleAlaTyrArgIleGluArgLeuGl 533
         :::   :::|||:::   ||||||:::      ||||||   ||||||||
     133 GATTACAACTCAACTCAGTTAAAACTTAAACAAAGAATAACAAGGTTAGA 84

     533 uTrpGlnThrArgGluTrpProMetProGluAspLeuAlaGluAlaAspA 550
         |   :::   ::::::
      83 AAGTAGAATAAAGAAA..................................... 68

     550 spGluGlyGluProGluGluGlyThrSerLysAspGluGluGlyArgAla 566
                       :::::|||   :::   :::|||   |||   :::
      67 .............TCCAATGGTGACACTGTATCTGAAGGTGGAGACAGT 32

     567 SerGlnGluGlyProPro 572
         :::   |||      |||
      31 GCTGCGGAAGAAGACCCT 14
```

```
OM of: US-09-927-267-1 to: GenEmbl:*   out_format : pfs

Date: Jun 20, 2002  6:29 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09927267/runat_19062002_090142_25489/app_query.fasta_1.640
 -DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500
 -MINLEN=0 -MAXLEN=2000000000 -USER=US09927267_@CGN1_1_8051
 -NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-927-267-1
Query length: 575
Database: GenEmbl:*
Database sequences: 1797656
Database length: 1873333701
Search time (sec): 2604.520000


score_list:
Sequence           Strd Orig    ZScore    EScore Len  ! Documentation  ..
gb_pat:AX355967      + 2984.00 4706.27  8.9e-254  2013  ! AX355967 Sequence 2 from Paten
gb_pat:AX350404      + 2984.00 4703.76  1.2e-253  2551  ! AX350404 Sequence 29 from Pate
gb_pat:AX350402      + 2957.50 4665.26  1.7e-251  1835  ! AX350402 Sequence 27 from Pate
gb_ro:RNU12425       + 2811.00 4432.56  1.6e-238  1938  ! U12425 Rattus norvegicus olfac
gb_ro:RNU12623       + 2811.00 4430.51  2.0e-238  2351  ! U12623 Rattus norvegicus Sprag
gb_htg:AC021935      + 2266.00 3521.97  8.2e-188  162965  ! AC021935 Homo sapiens chromc
gb_htg:AC022762      + 2266.00 3520.53  9.9e-188  186609  ! AC022762 Homo sapiens chromc
gb_htg:AC036216      + 2257.00 3507.61  5.2e-187  164434  ! AC036216 Homo sapiens chromc
gb_ro:MMU243933      + 1580.50 2479.02  1.0e-129  2781  ! AJ243933 Mus musculus mRNA for
gb_om:BTCNGCHAN      + 1579.00 2477.20  1.3e-129  2641  ! X89600 B.taurus mRNA for alpha
gb_om:BTCNG3         + 1579.00 2472.79  2.3e-129  3998  ! X76485 B.taurus CNG3 mRNA for
gb_ro:RNO272429      + 1578.00 2478.06  1.1e-129  2098  ! AJ272429 Rattus norvegicus mRN
gb_ro:RNO272428      + 1578.00 2477.49  1.2e-129  2212  ! AJ272428 Rattus norvegicus mRN
gb_ro:AB002801       + 1578.00 2477.49  1.2e-129  2213  ! AB002801 Rattus norvegicus mRN
gb_om:BTCAMPGC       + 1568.00 2457.84  1.5e-128  3166  ! X55010 B.taurus mRNA for cAMP-
gb_ov:GGCONECNG      + 1562.50 2450.83  3.8e-128  2697  ! X89598 G.gallus mRNA for alpha
gb_om:OCRACNG        + 1562.00 2450.15  4.1e-128  2669  ! X59668 O.cuniculus rACNG mRNA
gb_ro:AF126808       + 1558.00 2443.24  1.0e-127  2815  ! AF126808 Rattus norvegicus olf
gb_ro:RNOLFCH        + 1558.00 2442.47  1.1e-127  3027  ! X55519 Rat mRNA for a cyclic n
gb_pr:AF065314       + 1549.50 2427.51  7.5e-127  3486  ! AF065314 Homo sapiens cone pho
gb_om:BTCGMPCH       + 1544.50 2422.37  1.5e-126  2682  ! X51604 B.taurus RNA for cGMP-g
gb_ro:MMU19717       + 1543.00 2421.64  1.6e-126  2297  ! U19717 Mus musculus cGMP-gated
gb_om:CFACGMP        + 1541.50 2417.56  2.7e-126  2697  ! X99914 C.familiaris gene encod
gb_om:CFU83905       + 1541.50 2417.48  2.7e-126  2717  ! U83905 Canis familiaris rod ph
gb_om:SSU85404       + 1538.50 2415.66  3.4e-126  2061  ! U85404 Sus scrofa cGMP-gated c
gb_pr:S42457         + 1538.50 2412.19  5.4e-126  2857  ! S42457 CNCG=rod photoreceptor
gb_ro:RNU93851       + 1537.00 2412.84  4.9e-126  2148  ! U93851 Rattus norvegicus rod-t
gb_pat:AX179746      + 1535.00 2410.46  6.7e-126  1995  ! AX179746 Sequence 34 from Pate
gb_ro:MUSCNCG        + 1534.00 2406.59  1.1e-125  2474  ! M84742 Mus musculus cGMP-gated
gb_ro:RNU48803       + 1532.00 2405.40  1.3e-125  2052  ! U48803 Rattus norvegicus rod p
gb_ov:GGRODCNG       + 1529.00 2398.09  3.3e-125  2612  ! X89599 G.gallus mRNA for alpha
gb_pr:HUMCGMP        + 1526.50 2394.59  5.1e-125  2500  ! M84741 Human cGMP-gated cation
gb_ov:ICTCNC         + 1436.50 2254.10  3.4e-117  2049  ! M83111 Ictalurus punctatus cyc
gb_htg:AC091450      + 1399.00 2148.51  2.6e-111  158588  ! AC091450 Mus musculus chromc
gb_ro:MMU49391       + 1397.00 2175.10  8.6e-113  9618  ! U49391 Mus musculus cyclic nuc
gb_ov:AF393839       + 1380.50 2159.05  6.7e-112  3718  ! AF393839 Oncorhynchus mykiss c
gb_pr:HSAF002992     + 1355.00 2083.27  1.1e-107  104037  ! AF002992 Homo sapiens cosmid
gb_ro:MMU238241      + 1349.50 2120.09  1.0e-109  1428  ! AJ238241 Mus musculus CNG3 gen
gb_pr:AC092675       + 1344.00 2061.71  1.8e-106  153551  ! AC092675 Homo sapiens chromc
gb_htg:AL671880      + 1336.00 2046.23  1.3e-105  200039  ! AL671880 Mus musculus chromd
gb_htg:AC099698      - 1336.00 2045.27  1.5e-105  218872  ! AC099698 Mus musculus chr
em_htg_hum:AC013535  + 1323.00 2024.51  2.1e-104  222364  ! Ac013535 Homo sapiens chr
gb_htg:AC107068      - 1323.00 2024.16  2.2e-104  229687  ! AC107068 Homo sapiens chr
gb_in:DMCNGCHAN      + 1241.50 1941.39  9.0e-100  2913  ! X89601 D.melanogaster mRNA
gb_htg:AC007704      - 1210.50 1847.92  1.4e-94   190000  ! AC007704 Homo sapiens chr

seq_name: gb_pat:AX355967

seq_documentation_block:
LOCUS       AX355967               2013 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 2 from Patent WO0202633.
ACCESSION   AX355967
VERSION     AX355967.1  GI:18620561
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Raumann,B.E., Sanjanwala,M.S., Tribouley,C.M. and Walia,N.K.
  TITLE     Transporters and ion channels
  JOURNAL   Patent: WO 0202633-A 2 10-JAN-2002;
            Incyte Genomics, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .2013
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /note="Incyte ID No: 6703242CB1"
BASE COUNT      472 a    576 c    557 g    408 t
ORIGIN


alignment_scores:
           Quality: 2984.00              Length:    575
             Ratio:  5.199                 Gaps:      0
 Percent Similarity:  99.826   Percent Identity: 99.826

alignment_block:
 US-09-927-267-1 x AX355967   ..

 Align seg 1/1  to: AX355967  from: 1  to: 2013

       1 MetSerGlnAspThrLysValLysThrThrGluSerSerProProAlaPr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     124 ATGAGCCAGGACACCAAAGTGAAGACAACAGAGTCCAGTCCCCCAGCCCC 173

      17 oSerLysAlaArgLysLeuLeuProValLeuAspProSerGlyAspTyrT 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     174 ATCCAAGGCCAGGAAGTTGCTGCCTGTCCTGGACCCATCTGGGGATTACT 223

      34 yrTyrTrpTrpLeuAsnThrMetValPheProValMetTyrAsnLeuIle 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     224 ACTACTGGTGGCTGAACACAATGGTCTTCCCAGTCATGTATAACCTCATC 273

      51 IleLeuValCysArgAlaCysPheProAspLeuGlnHisGlyTyrLeuVa 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     274 ATCCTCGTGTGCAGAGCCTGCTTCCCCGACTTGCAGCACGGTTATCTGGT 323

      67 lAlaTrpLeuValLeuAspTyrThrSerAspLeuLeuTyrLeuLeuAspM 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     324 GGCCTGGTTGGTGCTGGACTACACGAGTGACCTGCTATACCTACTAGACA 373

      84 etValValArgPheHisThrGlyPheLeuGluGlnGlyIleLeuValVal 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     374 TGGTGGTGCGCTTCCACACAGGATTCTTGGAACAGGGCATCCTGGTGGTG 423

     101 AspLysGlyArgIleSerSerArgTyrValArgThrTrpSerPhePheLe 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     424 GACAAGGGTAGGATCTCGAGTCGCTACGTTCGCACCTGGAGTTTCTTCTT 473

     117 uAspLeuAlaSerLeuMetProThrAspValValTyrValArgLeuGlyP 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     474 GGACCTGGCTTCCCTGATGCCCACAGATGTGGTCTACGTGCGGCTGGGCC 523
```

```
  134 roHisThrProThrLeuArgLeuAsnArgPheLeuArgAlaProArgLeu 150
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  524 CGCACACACCCACCCTGAGGCTGAACCGCTTTCTCCGCGCGCCCCGCCTC 573

  151 PheGluAlaPheAspArgThrGluThrArgThrAlaTyrProAsnAlaPh 167
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  574 TTCGAGGCCTTCGACCGCACAGAGACCCGCACAGCTTACCCAAATGCCTT 623

  167 eArgIleAlaLysLeuMetLeuTyrIlePheValValIleHisTrpAsnS 184
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  624 TCGCATTGCCAAGCTGATGCTTTACATTTTTGTCGTCATCCATTGGAACA 673

  184 erCysLeuTyrPheAlaLeuSerArgTyrLeuGlyPheGlyArgAspAla 200
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  674 GCTGCCTATACTTTGCCCTATCCCGGTACCTGGGCTTCGGGCGTGACGCA 723

  201 TrpValTyrProAspProAlaGlnProGlyPheGluArgLeuArgArgGl 217
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  724 TGGGTGTACCCGGACCCCGCGCAGCCTGGCTTTGAGCGCCTGCGGCGCCA 773

  217 nTyrLeuTyrSerPheTyrPheSerThrLeuIleLeuThrThrValGlyA 234
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  774 GTACCTCTATAGCTTTTACTTCCCCACGCTGATACTGACTACAGTGGGCG 823

  234 spThrProProProAlaArgGluGluGluTyrLeuPheMetValGlyAsp 250
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  824 ATACACCGCCGCCAGCCAGGGAAGAAGAGTACCTCTTCATGGTGGGCGAC 873

  251 PheLeuLeuAlaValMetGlyPheAlaThrIleMetGlySerMetSerSe 267
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  874 TTCCTGCTGGCCGTCATGGGTTTCGCCACCATCATGGGTAGCATGAGCTC 923

  267 rValIleTyrAsnMetAsnThrAlaAspAlaAlaPheTyrProAspHisA 284
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  924 TGTCATCTACAACATGAACACTGCAGATGCGGCTTTCTACCCAGATCATG 973

  284 laLeuValLysLysTyrMetLysLeuGlnHisValAsnArgLysLeuGlu 300
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  974 CACTGGTGAAGAAGTACATGAAGCTGCAGCACGTCAACCGCAAGCTGGAG 1023

  301 ArgArgValIleAspTrpTyrGlnHisLeuGlnIleAsnLysLysMetTh 317
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1024 CGGCGAGTTATTGACTGGTATCAGCACCTGCAGATCAACAAGAAGATGAC 1073

  317 rAsnGluValAlaIleLeuGlnHisLeuProGluArgLeuArgAlaGluV 334
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1074 CAACGAGGTAGCCATCTTACAGCACTTGCCTGAGCGGCTGCGGGCAGAAG 1123

  334 alAlaValSerValHisLeuSerThrLeuSerArgValGlnIlePheGln 350
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1124 TGGCTGTGTCTGTGCACCTGTCCACTCTGAGCCGGGTGCAGATCTTTCAG 1173

  351 AsnCysGluAlaSerLeuLeuGluGluLeuValLeuLysLeuGlnProGl 367
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1174 AACTGTGAGGCCAGCCTGCTGGAGGAGCTGGTGCTGAAGCTGCAGCCCCA 1223

  367 nThrTyrSerProGlyGluTyrValCysArgLysGlyAspIleGlyGlnG 384
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1224 GACCTACTCACCAGGTGAATATGTATGCCGCAAAGGAGACATTGGCCAAG 1273

  384 luMetTyrIleIleArgGluGlyGlnLeuAlaValValAlaAspAspGly 400
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1274 AGATGTACATCATCCGAGAGGGTCAACTGGCCGTGGTGGCAGATGATGGT 1323

  401 IleThrGlnTyrAlaValLeuGlyAlaGlyLeuTyrPheGlyGluIleSe 417
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1324 ATCACACAGTATGCTGTGCTCGGTGCAGGGCTCTACTTTGGGGAGATCAG 1373

  417 rIleIleAsnIleLysGlyAsnMetSerGlyAsnArgArgThrAlaAsnI 434
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1374 CATCATCAACATCAAAGGGAACATGTCTGGGAACCGCCGCACAGCCAACA 1423

  434 leLysSerLeuGlyTyrSerAspLeuPheCysLeuSerLysGluAspLeu 450
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1424 TCAAGAGCCTAGGTTATTCAGACCTATTCTGCCTGAGCAAGGAGGACCTG 1473

  451 ArgGluValLeuSerGluTyrProGlnAlaGlnThrIleMetGluGluLy 467
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1474 CGGGAGGTGCTGAGCGAGTATCCACAAGCACAGACCATCATGGAGGAGAA 1523

  467 sGlyArgGluIleLeuLeuLysMetAsnLysLeuAspValAsnAlaGluA 484
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1524 AGGACGTGAGATCCTGCTGAAAATGAACAAGTTGGACGTGAATGCTGAGG 1573

  484 laAlaGluIleAlaLeuGlnGluAlaThrGluSerArgLeuArgGlyLeu 500
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1574 CAGCTGAGATCGCCCTGCAGGAGGCCACAGAGTCCCGGCTACGAGGCCTA 1623

  501 AspGlnGlnLeuAspAspLeuGlnThrLysPheAlaArgLeuLeuAlaGl 517
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1624 GACCAGCAGCTGGATGATCTACAGACCAAGTTTGCTCGCCTCCTGGCTGA 1673

  517 uLeuGluSerSerAlaLeuLysIleAlaTyrArgIleGluArgLeuGluT 534
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1674 GCTGGAGTCCAGCGCACTTAAGATTGCTTACCGCATTGAACGGCTGGAGT 1723

  534 rpGlnThrArgGluTrpProMetProGluAspLeuAlaGluAlaAspAsp 550
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1724 GGCAGACTCGAGAGTGGCCAATGCCCGAGGACCTGGCTGAGGCTGATGAC 1773

  551 GluGlyGluProGluGluGlyThrSerLysAspGluGluGlyArgAlaSe 567
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1774 GAGGGTGAGCCTGAGGAGGGAACTTCCAAAGATGAAGAGGGCAGGGCCAG 1823

  567 rGlnGluGlyProProGlyProGlu 575
      |||||||||||||||||||||||||
 1824 CCAGGAGGGACCCCCAGGTCCAGAG 1848
```

```
seq_name: gb_pat:AX350404

seq_documentation_block:
LOCUS       AX350404               2551 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 29 from Patent WO0181578.
ACCESSION   AX350404
VERSION     AX350404.1  GI:18616046
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Vernet,C.A., Fernandes,E.R., Gerlach,V., Shimkets,R.A.,
            Malyankar,U.M., Boldog,F.L., Zerhusen,B.D., Spytek,K.A.,
            Majumder,K., Tchernev,V.T., Padigaru,M., Patturajan,M.,
            Burgess,C.E., Gangolli,E.A., Smithson,G., Rastelli,L.,
            Macdougall,J.R., Taupier,R.J., Grosse,W.M. and Alsobrook,J.P.
  TITLE     Novel proteins and nucleic acids encoding same
  JOURNAL   Patent: WO 0181578-A 29 01-NOV-2001;
            Curagen Corporation (US)
FEATURES             Location/Qualifiers
     source          1. .2551
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      646 a    668 c    707 g    530 t
ORIGIN


alignment_scores:
            Quality: 2984.00             Length:    575
              Ratio:  5.190                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 99.826

alignment_block:
 US-09-927-267-1 x AX350404   ..
```

Align seg 1/1 to: AX350404 from: 1 to: 2551

```
      1 MetSerGlnAspThrLysValLysThrThrGluSerSerProProAlaPr 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    779 ATGAGCCAGGACACCAAAGTGAAGACAACAGAGTCCAGTCCCCCAGCCCC 828

     17 oSerLysAlaArgLysLeuLeuProValLeuAspProSerGlyAspTyrT 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    829 ATCCAAGGCCAGGAAGTTGCTGCCTGTCCTGGACCCATCTGGGGATTACT 878

     34 yrTyrTrpTrpLeuAsnThrMetValPheProValMetTyrAsnLeuIle 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    879 ACTACTGGTGGCTGAACACAATGGTCTTCCCAGTCATGTATAACCTCATC 928

     51 IleLeuValCysArgAlaCysPheProAspLeuGlnHisGlyTyrLeuVa 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    929 ATCCTCGTGTGCAGAGCCTGCTTCCCCGACTTGCAGCACGGTTATCTGGT 978

     67 lAlaTrpLeuValLeuAspTyrThrSerAspLeuLeuTyrLeuLeuAspM 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    979 GGCCTGGTTGGTGCTGGACTACACGAGTGACCTGCTATACCTACTAGACA 1028

     84 etValValArgPheHisThrGlyPheLeuGluGlnGlyIleLeuValVal 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1029 TGGTGGTGCGCTTCCACACAGGATTCTTGGAACAGGGCATCCTGGTGGTG 1078

    101 AspLysGlyArgIleSerSerArgTyrValArgThrTrpSerPhePheLe 117
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1079 GACAAGGGTAGGATCTCGAGTCGCTACGTTCGCACCTGGAGTTTCTTCTT 1128

    117 uAspLeuAlaSerLeuMetProThrAspValValTyrValArgLeuGlyP 134
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1129 GGACCTGGCTTCCCTGATGCCCACAGATGTGGTCTACGTGCGGCTGGGCC 1178

    134 roHisThrProThrLeuArgLeuAsnArgPheLeuArgAlaProArgLeu 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1179 CGCACACACCCACCCTGAGGCTGAACCGCTTTCTCCGCGCGCCCCGCCTC 1228

    151 PheGluAlaPheAspArgThrGluThrArgThrAlaTyrProAsnAlaPh 167
        ||||||||||||||||||:::||||||||||||||||||||||||||||
   1229 TTCGAGGCCTTCGACCGCGCAGAGACCCGCACAGCTTACCCAAATGCCTT 1278

    167 eArgIleAlaLysLeuMetLeuTyrIlePheValValIleHisTrpAsnS 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1279 TCGCATTGCCAAGCTGATGCTTTACATTTTTGTCGTCATCCATTGGAACA 1328

    184 erCysLeuTyrPheAlaLeuSerArgTyrLeuGlyPheGlyArgAspAla 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1329 GCTGCCTATACTTTGCCCTATCCCGGTACCTGGGCTTCGGGCGTGACGCA 1378

    201 TrpValTyrProAspProAlaGlnProGlyPheGluArgLeuArgArgGl 217
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1379 TGGGTGTACCCGGACCCCGCGCAGCCTGGCTTTGAGCGCCTGCGGCGCCA 1428

    217 nTyrLeuTyrSerPheTyrPheSerThrLeuIleLeuThrThrValGlyA 234
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1429 GTACCTCTATAGCTTTTACTTCTCCACGCTGATACTGACTACAGTGGGCG 1478

    234 spThrProProProAlaArgGluGluGluTyrLeuPheMetValGlyAsp 250
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1479 ATACACCGCCGCCAGCCAGGGAAGAAGAGTACCTCTTCATGGTGGGCGAC 1528

    251 PheLeuLeuAlaValMetGlyPheAlaThrIleMetGlySerMetSerSe 267
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1529 TTCCTGCTGGCCGTCATGGGTTTCGCCACCATCATGGGTAGCATGAGCTC 1578

    267 rValIleTyrAsnMetAsnThrAlaAspAlaAlaPheTyrProAspHisA 284
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1579 TGTCATCTACAACATGAACACTGCAGATGCGGCTTTCTACCCAGATCATG 1628

    284 laLeuValLysLysTyrMetLysLeuGlnHisValAsnArgLysLeuGlu 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1629 CACTGGTGAAGAAGTACATGAAGCTGCAGCACGTCAACCGCAAGCTGGAG 1678

    301 ArgArgValIleAspTrpTyrGlnHisLeuGlnIleAsnLysLysMetTh 317
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1679 CGGCGAGTTATTGACTGGTATCAGCACCTGCAGATCAACAAGAAGATGAC 1728

    317 rAsnGluValAlaIleLeuGlnHisLeuProGluArgLeuArgAlaGluV 334
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1729 CAACGAGGTAGCCATCTTACAGCACTTGCCTGAGCGGCTGCGGGCAGAAG 1778

    334 alAlaValSerValHisLeuSerThrLeuSerArgValGlnIlePheGln 350
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1779 TGGCTGTGTCTGTGCACCTGTCCACTCTGAGCCGGGTGCAGATCTTTCAG 1828

    351 AsnCysGluAlaSerLeuLeuGluGluLeuValLeuLysLeuGlnProGl 367
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1829 AACTGTGAGGCCAGCCTGCTGGAGGAGCTGGTGCTGAAGCTGCAGCCCCA 1878

    367 nThrTyrSerProGlyGluTyrValCysArgLysGlyAspIleGlyGlnG 384
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1879 GACCTACTCACCAGGTGAATATGTATGCCGCAAAGGAGACATTGGCCAAG 1928

    384 luMetTyrIleIleArgGluGlyGlnLeuAlaValValAlaAspAspGly 400
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1929 AGATGTACATCATCCGAGAGGGTCAACTGGCCGTGGTGGCAGATGATGGT 1978

    401 IleThrGlnTyrAlaValLeuGlyAlaGlyLeuTyrPheGlyGluIleSe 417
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1979 ATCACACAGTATGCTGTGCTCGGTGCAGGGCTCTACTTTGGGGAGATCAG 2028

    417 rIleIleAsnIleLysGlyAsnMetSerGlyAsnArgArgThrAlaAsnI 434
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2029 CATCATCAACATCAAAGGGAACATGTCTGGGAACCGCCGCACAGCCAACA 2078

    434 leLysSerLeuGlyTyrSerAspLeuPheCysLeuSerLysGluAspLeu 450
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2079 TCAAGAGCCTAGGTTATTCAGACCTATTCTGCCTGAGCAAGGAGGACCTG 2128

    451 ArgGluValLeuSerGluTyrProGlnAlaGlnThrIleMetGluGluLy 467
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2129 CGGGAGGTGCTGAGCGAGTATCCACAAGCACAGACCATCATGGAGGAGAA 2178

    467 sGlyArgGluIleLeuLeuLysMetAsnLysLeuAspValAsnAlaGluA 484
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2179 AGGACGTGAGATCCTGCTGAAAATGAACAAGTTGGACGTGAATGCTGAGG 2228

    484 laAlaGluIleAlaLeuGlnGluAlaThrGluSerArgLeuArgGlyLeu 500
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2229 CAGCTGAGATCGCCCTGCAGGAGGCCACAGAGTCCCGGCTACGAGGCCTA 2278

    501 AspGlnGlnLeuAspAspLeuGlnThrLysPheAlaArgLeuLeuAlaGl 517
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2279 GACCAGCAGCTGGATGATCTACAGACCAAGTTTGCTCGCCTCCTGGCTGA 2328

    517 uLeuGluSerSerAlaLeuLysIleAlaTyrArgIleGluArgLeuGluT 534
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2329 GCTGGAGTCCAGCGCACTTAAGATTGCTTACCGCATTGAACGGCTGGAGT 2378

    534 rpGlnThrArgGluTrpProMetProGluAspLeuAlaGluAlaAspAsp 550
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2379 GGCAGACTCGAGAGTGGCCAATGCCCGAGGACCTGGCTGAGGCTGATGAC 2428

    551 GluGlyGluProGluGluGlyThrSerLysAspGluGluGlyArgAlaSe 567
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2429 GAGGGTGAGCCTGAGGAGGGAACTTCCAAAGATGAAGAGGGCAGGGCCAG 2478

    567 rGlnGluGlyProProGlyProGlu 575
        ||||||||||||||||||||||||
   2479 CCAGGAGGGACCCCCAGGTCCAGAG 2503
```

seq_name: gb_pat:AX350402

seq_documentation_block:
LOCUS AX350402 1835 bp DNA linear PAT 06-FEB-2002

```
DEFINITION  Sequence 27 from Patent WO0181578.
ACCESSION   AX350402
VERSION     AX350402.1  GI:18616045
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Vernet,C.A., Fernandes,E.R., Gerlach,V., Shimkets,R.A.,
            Malyankar,U.M., Boldog,F.L., Zerhusen,B.D., Spytek,K.A.,
            Majumder,K., Tchernev,V.T., Padigaru,M., Patturajan,M.,
            Burgess,C.E., Gangolli,E.A., Smithson,G., Rastelli,L.,
            Macdougall,J.R., Taupier,R.J., Grosse,W.M. and Alsobrook,J.P.
  TITLE     Novel proteins and nucleic acids encoding same
  JOURNAL   Patent: WO 0181578-A 27 01-NOV-2001;
            Curagen Corporation (US)
FEATURES             Location/Qualifiers
     source          1. .1835
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      420 a    521 c    521 g    373 t
ORIGIN


alignment_scores:
            Quality: 2957.50             Length:    578
              Ratio:  5.143                Gaps:      3
 Percent Similarity:  99.481   Percent Identity: 99.481

alignment_block:
 US-09-927-267-1 x AX350402   ..

 Align seg 1/1  to: AX350402  from: 1  to: 1835

      1 MetSerGlnAspThrLysValLysThrThrGluSerSerProProAlaPr 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     54 ATGAGCCAGGACACCAAAGTGAAGACAACAGAGTCCAGTCCCCCAGCCCC 103

     17 oSerLysAla...ArgLysLeuLeuProValLeuAspProSerGlyAspT 33
        ||||||||||   ||||||||||||||||||||||||||||||||||||||
    104 ATCCAAGGCCAGGAGGAAGTTGCTGCCTGTCCTGGACCCATCTGGGGATT 153

     33 yrTyrTyrTrpTrpLeuAsnThrMetValPheProValMetTyrAsnLeu 49
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    154 ACTACTACTGGTGGCTGAACACAATGGTCTTCCCAGTCATGTATAACCTC 203

     50 IleIleLeuValCysArgAlaCysPheProAspLeuGlnHisGlyTyrLe 66
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    204 ATCATCCTCGTGTGCAGGGCCTGCTTCCCCGACTTGCAGCACGGTTATCT 253

     66 uValAlaTrpLeuValLeuAspTyrThrSerAspLeuLeuTyrLeuLeuA 83
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    254 GGTGGCCTGGTTGGTGCTGGACTACACGAGTGACCTGCTATACCTACTAG 303

     83 spMetValValArgPheHisThr...GlyPheLeuGluGlnGlyIleLeu 98
        ||||||||||||||||||||||   ||||||||||||||||||||||||
    304 ACATGGTGGTGCGCTTCCACACAGGTGGATTCTTGGAACAGGGCATCCTG 353

     99 ValValAspLysGlyArgIleSerSerArgTyrValArgThrTrpSerPh 115
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    354 GTGGTGGACAAGGGTAGGATCTCGAGTCGCTACGTTCGCACCTGGAGTTT 403

    115 ePheLeuAspLeuAlaSerLeuMetProThrAspValValTyrValArgL 132
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    404 CTTCTTGGACCTGGCTTCCCTGATGCCCACAGATGTGGTCTACGTGCGGC 453

    132 euGlyProHisThrProThrLeuArgLeuAsnArgPheLeuArgAlaPro 148
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    454 TGGGCCCGCACACACCCACCCTGAGGCTGAACCGCTTTCTCCGCGCGCCC 503

    149 ArgLeuPheGluAlaPheAspArgThrGluThrArgThrAlaTyrProAs 165
        ||||||||||||||||||||||||||||||||||||||||||||||||||

    504 CGCCTCTTCGAGGCCTTCGACCGCACAGAGACCCGCACAGCTTACCCAAA 553

    165 nAlaPheArgIleAlaLysLeuMetLeuTyrIlePheValValIleHisT 182
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    554 TGCCTTTCGCATTGCCAAGCTGATGCTTTACATTTTTGTCGTCATCCATT 603

    182 rpAsnSerCysLeuTyrPheAlaLeuSerArgTyrLeuGlyPheGlyArg 198
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    604 GGAACAGCTGCCTATACTTTGCCCTATCCCGGTACCTGGGCTTCGGGCGT 653

    199 AspAlaTrpValTyrProAspProAlaGlnProGlyPheGluArgLeuAr 215
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    654 GACGCATGGGTGTACCCGGACCCCGCGCAGCCTGGCTTTGAGCGCCTGCG 703

    215 gArgGlnTyrLeuTyrSerPheTyrPheSerThrLeuIleLeuThrThrV 232
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    704 GCGCCAGTACCTCTATAGCTTTTACTTCTCCACGCTGATACTGACTACAG 753

    232 alGlyAspThrProProProAlaArgGluGluGluTyrLeuPheMetVal 248
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    754 TGGGCGATACACCGCCGCCAGCCAGGGAAGAAGAGTACCTCTTCATGGTG 803

    249 GlyAspPheLeuLeuAlaValMetGlyPheAlaThrIleMetGlySerMe 265
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    804 GGCGACTTCCTGCTGGCCGTCATGGGTTTCGCCACCATCATGGGTAGCAT 853

    265 tSerSerValIleTyrAsnMetAsnThrAlaAspAlaAlaPheTyrProA 282
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    854 GAGCTCTGTCATCTACAACATGAACACTGCAGATGCGGCTTTCTACCCAG 903

    282 spHisAlaLeuValLysLysTyrMetLysLeuGlnHisValAsnArgLys 298
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    904 ATCATGCACTGGTGAAGAAGTACATGAAGCTGCAGCACGTCAACCGCAAG 953

    299 LeuGluArgArgValIleAspTrpTyrGlnHisLeuGlnIleAsnLysLy 315
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    954 CTGGAGCGGCGAGTTATTGACTGGTATCAGCACCTGCAGATCAACAAGAA 1003

    315 sMetThrAsnGluValAlaIleLeuGlnHisLeuProGluArgLeuArgA 332
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1004 GATGACCAACGAGGTAGCCATCTTACAGCACTTGCCTGAGCGGCTGCGGG 1053

    332 laGluValAlaValSerValHisLeuSerThrLeuSerArgValGlnIle 348
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1054 CAGAAGTGGCTGTGTCTGTGCACCTGTCCACTCTGAGCCGGGTGCAGATC 1103

    349 PheGlnAsnCysGluAlaSerLeuLeuGluGluLeuValLeuLysLeuGl 365
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1104 TTTCAGAACTGTGAGGCCAGCCTGCTGGAGGAGCTGGTGCTGAAGCTGCA 1153

    365 nProGlnThrTyrSerProGlyGluTyrValCysArgLysGlyAspIleG 382
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1154 GCCCCAGACCTACTCACCAGGTGAATATGTATGCCGCAAAGGAGACATTG 1203

    382 lyGlnGluMetTyrIleIleArgGluGlyGlnLeuAlaValValAlaAsp 398
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1204 GCCAAGAGATGTACATCATCCGAGAGGGTCAACTGGCCGTGGTGGCAGAT 1253

    399 AspGlyIleThrGlnTyrAlaValLeuGlyAlaGlyLeuTyrPheGlyGl 415
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1254 GATGGTATCACACAGTATGCTGTGCTCGGTGCAGGGCTCTACTTTGGGGA 1303

    415 uIleSerIleIleAsnIleLys...GlyAsnMetSerGlyAsnArgArgT 431
        ||||||||||||||||||||||   ||||||||||||||||||||||||
   1304 GATCAGCATCATCAACATCAAAGGTGGGAACATGTCTGGGAACCGCCGCA 1353

    431 hrAlaAsnIleLysSerLeuGlyTyrSerAspLeuPheCysLeuSerLys 447
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1354 CAGCCAACATCAAGAGCCTAGGTTATTCAGACCTATTCTGCCTGAGCAAG 1403

    448 GluAspLeuArgGluValLeuSerGluTyrProGlnAlaGlnThrIleMe 464
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1404 GAGGACCTGCGGGAGGTGCTGAGCGAGTATCCACAAGCACAGACCATCAT 1453
```

```
     464 tGluGluLysGlyArgGluIleLeuLeuLysMetAsnLysLeuAspValA 481
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1454 GGAGGAGAAAGGACGTGAGATCCTGCTGAAAATGAACAAGTTGGACGTGA 1503

     481 snAlaGluAlaAlaGluIleAlaLeuGlnGluAlaThrGluSerArgLeu 497
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1504 ATGCTGAGGCAGCTGAGATCGCCCTGCAGGAGGCCACAGAGTCCCGGCTA 1553

     498 ArgGlyLeuAspGlnGlnLeuAspAspLeuGlnThrLysPheAlaArgLe 514
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1554 CGAGGCCTAGACCAGCAGCTGGATGATCTACAGACCAAGTTTGCTCGCCT 1603

     514 uLeuAlaGluLeuGluSerSerAlaLeuLysIleAlaTyrArgIleGluA 531
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1604 CCTGGCTGAGCTGGAGTCCAGCGCACTTAAGATTGCTTACCGCATTGAAC 1653

     531 rgLeuGluTrpGlnThrArgGluTrpProMetProGluAspLeuAlaGlu 547
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1654 GGCTGGAGTGGCAGACTCGAGAGTGGCCAATGCCCGAGGACCTGGCTGAG 1703

     548 AlaAspAspGluGlyGluProGluGluGlyThrSerLysAspGluGluGl 564
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1704 GCTGATGACGAGGGTGAGCCTGAGGAGGGAACTTCCAAAGATGAAGAGGG 1753

     564 yArgAlaSerGlnGluGlyProProGlyProGlu 575
         ||||||||||||||||||||||||||||||||
    1754 CAGGGCCAGCCAGGAGGGACCCCCAGGTCCAGAG 1787

seq_name: gb_ro:RNU12425

seq_documentation_block:
LOCUS       RNU12425                1938 bp    mRNA    linear   ROD 05-APR-1995
DEFINITION  Rattus norvegicus olfactory cyclic nucleotide-gated channel mRNA,
            complete cds.
ACCESSION   U12425
VERSION     U12425.1  GI:548083
KEYWORDS    .
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1938)
  AUTHORS   Liman,E.R. and Buck,L.B.
  TITLE     A second subunit of the olfactory cyclic nucleotide-gated channel
            confers high sensitivity to cAMP
  JOURNAL   Neuron 13 (3), 611-621 (1994)
  MEDLINE   95000663
REFERENCE   2  (bases 1 to 1938)
  AUTHORS   Liman,E.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-1994) Emily R. Liman, Neurobiology, Harvard
            Medical School, 220 Longwood Avenue, Boston, MA 02115, USA
FEATURES             Location/Qualifiers
     source          1. .1938
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /tissue_type="olfactory neuroepithelium"
     mRNA            1. .1937
                     /evidence=experimental
     CDS             107. .1834
                     /codon_start=1
                     /product="olfactory cyclic nucleotide-gated channel"
                     /protein_id="AAA64748.1"
                     /db_xref="GI:548084"
                     /translation="MSQDGKVKTTESTPPAPTKARKWLPVLDPSGDYYYWWLNTMVFP
                     IMYNLIIVVCRACFPDLQHSYLVAWFVLDYTSDLLYLLDIGVRFHTGFLEQGILVVDK
                     GMIASRYVRTWSFLLDLASLVPTDAAYVQLGPHIPTLRLNRFLRVPRLFEAFDRTETR
                     TAYPNAFRIAKLMLYIFVVIHWNSCLYFALSRYLGFGRDAWVYPDPAQPGFERLRRQY
                     LYSFYFSTLILTTVGDTPLPDREEEYLFMVGDFLLAVMGFATIMGSMSSVIYNMNTAD
                     AAFYPDHALVKKYMKLQHVNKRLERRVIDWYQHLQINKKMTNEVAILQHLPERLRAEV
                     AVSVHLSTLSRVQIFQNCEASLLEELVLKLQPQTYSPGEYVCRKGDIGREMYIIREGQ
                     LAVVADDGVTQYAVLGAGLYFGEISIINIKGNMSGNRRTANIKSLGYSDLFCLSKEDL
                     REVLSEYPQAQAVMEEKGREILLKMNKLDVNAEAAEIALQEATESRLKGLDQQLDDLQ
                     TKFARLLAELESSALKIAYRIERLEWQTREWPMPEDMGEADDEAEPGEGTSKDGEGKA
                     GQAGPSGIE"
     polyA_site      1938
                     /note="3 A residues"
BASE COUNT      458 a    531 c    543 g    406 t
ORIGIN


alignment_scores:
           Quality: 2811.00            Length:    575
             Ratio:  4.993              Gaps:      0
 Percent Similarity:  97.913   Percent Identity: 93.391

alignment_block:
 US-09-927-267-1 x RNU12425   ..

 Align seg 1/1  to: RNU12425  from: 1  to: 1938

       1 MetSerGlnAspThrLysValLysThrThrGluSerSerProProAlaPr 17
         ||||||||||||   ||||||||||||||||||||:::||||||||||||
     107 ATGAGCCAGGACGGCAAAGTGAAGACCACAGAGTCCACGCCCCCAGCCCC 156

      17 oSerLysAlaArgLysLeuLeuProValLeuAspProSerGlyAspTyrT 34
         |:::||||||||||||   ||||||||||||||||||||||||||||||
     157 AACCAAAGCCAGGAAGTGGCTGCCAGTCCTAGACCCATCTGGGGATTACT 206

      34 yrTyrTrpTrpLeuAsnThrMetValPheProValMetTyrAsnLeuIle 50
         ||||||||||||||||||||||||||||||:::||||||||||||||||
     207 ACTACTGGTGGCTGAACACAATGGTCTTCCCAATCATGTACAACCTCATC 256

      51 IleLeuValCysArgAlaCysPheProAspLeuGlnHisGlyTyrLeuVa 67
         |||:::||||||||||||||||||||||||||||||||:::||||||||
     257 ATCGTTGTATGCAGGGCCTGCTTTCCTGACTTGCAGCACAGTTACCTGGT 306

      67 lAlaTrpLeuValLeuAspTyrThrSerAspLeuLeuTyrLeuLeuAspM 84
         |||||||:::||||||||||||||||||||||||||||||||||||||||:
     307 GGCCTGGTTCGTGCTGGACTACACGAGTGACCTGCTGTACCTACTAGATA 356

      84 etValValArgPheHisThrGlyPheLeuGluGlnGlyIleLeuValVal 100
         ::   ||||||||||||||||||||||||||||||||||||||||||||||
     357 TTGGGGTACGCTTCCACACAGGATTCCTAGAGCAGGGCATCCTGGTGGTA 406

     101 AspLysGlyArgIleSerSerArgTyrValArgThrTrpSerPhePheLe 117
         |||||||||   |||:::|||||||||||||||||||||||||||:::||
     407 GACAAAGGCATGATCGCCAGTCGCTACGTCCGCACCTGGAGCTTCCTGTT 456

     117 uAspLeuAlaSerLeuMetProThrAspValValTyrValArgLeuGlyP 134
         |||||||||||||||:::|||||||||:::::|||||:::|||||||||
     457 GGACCTGGCGTCCCTGGTCCCCACAGATGCGGCCTATGTGCAGCTGGGCC 506

     134 roHisThrProThrLeuArgLeuAsnArgPheLeuArgAlaProArgLeu 150
         |||||   ||||||||||||||||||||||||||||:::|||||||||
     507 CCCACATCCCTACACTCCGGCTAAACCGCTTTCTCCGAGTGCCCCGCCTC 556

     151 PheGluAlaPheAspArgThrGluThrArgThrAlaTyrProAsnAlaPh 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     557 TTCGAGGCTTTTGATCGTACAGAGACCCGCACGGCTTACCCAAATGCCTT 606

     167 eArgIleAlaLysLeuMetLeuTyrIlePheValValIleHisTrpAsnS 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     607 CCGCATAGCCAAGCTGATGCTTTACATTTTTGTTGTCATCCATTGGAACA 656

     184 erCysLeuTyrPheAlaLeuSerArgTyrLeuGlyPheGlyArgAspAla 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     657 GTTGCTTATACTTCGCCCTGTCCAGGTACCTGGGCTTTGGACGGGATGCG 706

     201 TrpValTyrProAspProAlaGlnProGlyPheGluArgLeuArgArgGl 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     707 TGGGTATACCCAGACCCTGCGCAACCTGGCTTTGAGCGCTTGCGGCGCCA 756

     217 nTyrLeuTyrSerPheTyrPheSerThrLeuIleLeuThrThrValGlyA 234
```

```
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 757 GTATCTCTACAGCTTCTACTTCTCCACTCTGATCCTGACCACAGTGGGTG 806

 234 spThrProProProAlaArgGluGluGluTyrLeuPheMetValGlyAsp 250
     ||||||||   |||   ||||||||||||||||||||||||||||||||||||||||||||||||
 807 ACACGCCGCTGCCAGACCGAGAGGAAGAGTACCTCTTCATGGTGGGTGAC 856

 251 PheLeuLeuAlaValMetGlyPheAlaThrIleMetGlySerMetSerSe 267
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 857 TTCCTGCTGGCCGTCATGGGTTTCGCCACCATCATGGGTAGCATGAGCTC 906

 267 rValIleTyrAsnMetAsnThrAlaAspAlaAlaPheTyrProAspHisA 284
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 907 TGTCATTTACAACATGAACACTGCAGATGCGGCCTTCTACCCAGACCATG 956

 284 laLeuValLysLysTyrMetLysLeuGlnHisValAsnArgLysLeuGlu 300
     ||||||||||||||||||||||||||||||||||||||||:::::|||||||
 957 CGCTGGTAAAGAAGTACATGAAGCTGCAGCATGTCAACAAGAGGCTGGAG 1006

 301 ArgArgValIleAspTrpTyrGlnHisLeuGlnIleAsnLysLysMetTh 317
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1007 CGGCGAGTTATTGACTGGTACCAGCATCTTCAGATCAACAAGAAGATGAC 1056

 317 rAsnGluValAlaIleLeuGlnHisLeuProGluArgLeuArgAlaGluV 334
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1057 CAACGAGGTAGCCATCTTGCAGCACCTGCCTGAGCGGCTGCGGGCGGAGG 1106

 334 alAlaValSerValHisLeuSerThrLeuSerArgValGlnIlePheGln 350
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1107 TTGCTGTGTCCGTGCACCTGTCTACCCTGAGCCGAGTACAGATCTTCCAG 1156

 351 AsnCysGluAlaSerLeuLeuGluGluLeuValLeuLysLeuGlnProGl 367
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1157 AACTGTGAAGCCAGCCTGCTGGAAGAGCTGGTGCTGAAGCTACAGCCCCA 1206

 367 nThrTyrSerProGlyGluTyrValCysArgLysGlyAspIleGlyGlnG 384
     ||||||||||||||||||||||||||||||||||||||||||||||||||||:::|
1207 GACCTACTCGCCAGGCGAATATGTGTGCCGCAAAGGGGACATTGGCCGAG 1256

 384 luMetTyrIleIleArgGluGlyGlnLeuAlaValValAlaAspAspGly 400
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1257 AGATGTACATCATCCGTGAGGGCCAGCTGGCTGTGGTGGCCGATGATGGT 1306

 401 IleThrGlnTyrAlaValLeuGlyAlaGlyLeuTyrPheGlyGluIleSe 417
     :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1307 GTCACACAGTATGCTGTGCTTGGTGCAGGGCTCTACTTTGGGGAGATCAG 1356

 417 rIleIleAsnIleLysGlyAsnMetSerGlyAsnArgArgThrAlaAsnI 434
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1357 TATCATCAACATCAAAGGGAACATGTCTGGAAACCGACGAACAGCCAACA 1406

 434 leLysSerLeuGlyTyrSerAspLeuPheCysLeuSerLysGluAspLeu 450
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1407 TCAAGAGCCTAGGTTATTCAGACCTGTTCTGCCTCAGCAAGGAGGATCTG 1456

 451 ArgGluValLeuSerGluTyrProGlnAlaGlnThrIleMetGluGluLy 467
     ||||||||||||||||||||||||||||||||:::::|||||||||||||||
1457 CGGGAGGTACTGAGTGAGTACCCACAGGCCCAGGCGGTCATGGAGGAGAA 1506

 467 sGlyArgGluIleLeuLeuLysMetAsnLysLeuAspValAsnAlaGluA 484
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1507 GGGCCGAGAAATCTTGCTCAAAATGAATAAGTTGGATGTGAATGCTGAGG 1556

 484 laAlaGluIleAlaLeuGlnGluAlaThrGluSerArgLeuArgGlyLeu 500
     ||||||||||||||||||||||||||||||||||||||||:::|||||||
1557 CAGCTGAGATCGCCCTCCAGGAGGCCACAGAGTCTCGGCTCAAAGGCCTC 1606

 501 AspGlnGlnLeuAspAspLeuGlnThrLysPheAlaArgLeuLeuAlaGl 517
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1607 GACCAGCAGCTTGATGATCTGCAGACCAAGTTTGCTCGCCTACTGGCTGA 1656

 517 uLeuGluSerSerAlaLeuLysIleAlaTyrArgIleGluArgLeuGluT 534
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
1657 GCTGGAGTCCAGTGCACTGAAGATAGCTTACCGCATCGAAAGGCTGGAGT 1706

 534 rpGlnThrArgGluTrpProMetProGluAspLeuAlaGluAlaAspAsp 550
     ||||||||||||||||||||||||||||||||||:::::|||||||||||
1707 GGCAGACTCGAGAGTGGCCAATGCCAGAGGACATGGGTGAGGCTGATGAT 1756

 551 GluGlyGluProGluGluGlyThrSerLysAspGluGluGlyArgAlaSe 567
     |||:::|||||||   |||||||||||||||||   ||||||:::|||::
1757 GAGGCTGAGCCTGGAGAAGGGACGTCCAAGGATGGAGAGGGAAAGGCTGG 1806

 567 rGlnGluGlyProProGlyProGlu 575
     :|||   ||||||   |||   |||
1807 CCAGGCGGGACCCTCAGGCATAGAA 1831
```

```
seq_name: gb_ro:RNU12623

seq_documentation_block:
LOCUS       RNU12623                2351 bp    mRNA    linear   ROD 15-SEP-1994
DEFINITION  Rattus norvegicus Sprague-Dawley cyclic nucleotide gated cation
            channel mRNA, complete cds.
ACCESSION   U12623
VERSION     U12623.1  GI:538128
KEYWORDS    .
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 2351)
  AUTHORS   Bradley,J., Li,J., Davidson,N., Lester,H.A. and Zinn,K.
  TITLE     Heteromeric olfactory cyclic nucleotide-gated channels: A new
            subunit that confers increased sensitivity to cAMP
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 91, 8890-8894 (1994)
  MEDLINE   94377458
REFERENCE   2  (bases 1 to 2351)
  AUTHORS   Bradley,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-1994) Jonathan Bradley, Biology, California
            Institute of Technology, Pasadena, CA 91125, USA
COMMENT     On Sep 15, 1994 this sequence version replaced gi:521147.
FEATURES             Location/Qualifiers
     source          1. .2351
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
     CDS             109. .1836
                     /codon_start=1
                     /product="cyclic nucleotide gated cation channel"
                     /protein_id="AAA21464.1"
                     /db_xref="GI:538129"
                     /translation="MSQDGKVKTTESTPPAPTKARKWLPVLDPSGDYYYWWLNTMVFP
                     IMYNLIIVVCRACFPDLQHSYLVAWFVLDYTSDLLYLLDIGVRFHTGFLEQGILVVDK
                     GMIASRYVRTWSFLLDLASLVPTDAAYVQLGPHIPTLRLNRFLRVPRLFEAFDRTETR
                     TAYPNAFRIAKLMLYIFVVIHWNSCLYFALSRYLGFGRDAWVYPDPAQPGFERLRRQY
                     LYSFYFSTLILTTVGDTPLPDREEEYLFMVGDFLLAVMGFATIMGSMSSVIYNMNTAD
                     AAFYPDHALVKKYMKLQHVNKRLERRVIDWYQHLQINKKMTNEVAILQHLPERLRAEV
                     AVSVHLSTLSRVQIFQNCEASLLEELVLKLQPQTYSPGEYVCRKGDIGREMYIIREGQ
                     LAVVADDGVTQYAVLGAGLYFGEISIINIKGNMSGNRRTANIKSLGYSDLFCLSKEDL
                     REVLSEYPQAQAVMEEKGREILLKMNKLDVNAEAAEIALQEATESRLKGLDQQLDDLQ
                     TKFARLLAELESSALKIAYRIERLEWQTREWPMPEDMGEADDEAEPGEGTSKDGEGKA
                     GQAGPSGIE"
BASE COUNT      580 a    624 c    637 g    510 t
ORIGIN


alignment_scores:
            Quality: 2811.00             Length:    575
              Ratio:  4.993               Gaps:      0
 Percent Similarity:  97.913   Percent Identity: 93.391

alignment_block:
 US-09-927-267-1 x RNU12623   ..
```

Align seg 1/1 to: RNU12623 from: 1 to: 2351

```
      1 MetSerGlnAspThrLysValLysThrThrGluSerSerProProAlaPr 17
        ||||||||||||   ||||||||||||||||||||:::||||||||||||
    109 ATGAGCCAGGACGGCAAAGTGAAGACCACAGAGTCCACGCCCCCAGCCCC 158

     17 oSerLysAlaArgLysLeuLeuProValLeuAspProSerGlyAspTyrT 34
        |:::|||||||||||   ||||||||||||||||||||||||||||||||
    159 AACCAAAGCCAGGAAGTGGCTGCCAGTCCTAGACCCATCTGGGGATTACT 208

     34 yrTyrTrpTrpLeuAsnThrMetValPheProValMetTyrAsnLeuIle 50
        ||||||||||||||||||||||||||||||:::||||||||||||||||
    209 ACTACTGGTGGCTGAACACAATGGTCTTCCCAATCATGTACAACCTCATC 258

     51 IleLeuValCysArgAlaCysPheProAspLeuGlnHisGlyTyrLeuVa 67
        |||:::|||||||||||||||||||||||||||||||:::||||||||
    259 ATCGTTGTATGCAGGGCCTGCTTTCCTGACTTGCAGCACAGTTACCTGGT 308

     67 lAlaTrpLeuValLeuAspTyrThrSerAspLeuLeuTyrLeuLeuAspM 84
        |||||||:::|||||||||||||||||||||||||||||||||||||:
    309 GGCCTGGTTCGTGCTGGACTACACGAGTGACCTGCTGTACCTACTAGATA 358

     84 etValValArgPheHisThrGlyPheLeuGluGlnGlyIleLeuValVal 100
        ::  ||||||||||||||||||||||||||||||||||||||||||||||
    359 TTGGGGTACGCTTCCACACAGGATTCCTAGAGCAGGGCATCCTGGTGGTA 408

    101 AspLysGlyArgIleSerSerArgTyrValArgThrTrpSerPhePheLe 117
        |||||||||   |||:::|||||||||||||||||||||||||:::||
    409 GACAAAGGCATGATCGCCAGTCGCTACGTCCGCACCTGGAGCTTCCTGTT 458

    117 uAspLeuAlaSerLeuMetProThrAspValValTyrValArgLeuGlyP 134
        |||||||||||||||:::|||||||||:::::|||||||:::|||||||
    459 GGACCTGGCGTCCCTGGTCCCCACAGATGCGGCCTATGTGCAGCTGGGCC 508

    134 roHisThrProThrLeuArgLeuAsnArgPheLeuArgAlaProArgLeu 150
        |||||   ||||||||||||||||||||||||||||:::|||||||||
    509 CCCACATCCCTACACTCCGGCTAAACCGCTTTCTCCGAGTGCCCCGCCTC 558

    151 PheGluAlaPheAspArgThrGluThrArgThrAlaTyrProAsnAlaPh 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    559 TTCGAGGCTTTTGATCGTACAGAGACCCGCACGGCTTACCCAAATGCCTT 608

    167 eArgIleAlaLysLeuMetLeuTyrIlePheValValIleHisTrpAsnS 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    609 CCGCATAGCCAAGCTGATGCTTTACATTTTTGTTGTCATCCATTGGAACA 658

    184 erCysLeuTyrPheAlaLeuSerArgTyrLeuGlyPheGlyArgAspAla 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    659 GTTGCTTATACTTCGCCCTGTCCAGGTACCTGGGCTTTGGACGGGATGCG 708

    201 TrpValTyrProAspProAlaGlnProGlyPheGluArgLeuArgArgGl 217
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    709 TGGGTATACCCAGACCCTGCGCAACCTGGCTTTGAGCGCTTGCGGCGCCA 758

    217 nTyrLeuTyrSerPheTyrPheSerThrLeuIleLeuThrThrValGlyA 234
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    759 GTATCTCTACAGCTTCTACTTCTCCACTCTGATCCTGACCACAGTGGGTG 808

    234 spThrProProProAlaArgGluGluGluTyrLeuPheMetValGlyAsp 250
        ||||||||   |||   ||||||||||||||||||||||||||||||||
    809 ACACGCCGCTGCCAGACCGAGAGGAAGAGTACCTCTTCATGGTGGGTGAC 858

    251 PheLeuLeuAlaValMetGlyPheAlaThrIleMetGlySerMetSerSe 267
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    859 TTCCTGCTGGCCGTCATGGGTTTCGCCACCATCATGGGTAGCATGAGCTC 908

    267 rValIleTyrAsnMetAsnThrAlaAspAlaAlaPheTyrProAspHisA 284
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    909 TGTCATTTACAACATGAACACTGCAGATGCGGCCTTCTACCCAGACCATG 958

    284 laLeuValLysLysTyrMetLysLeuGlnHisValAsnArgLysLeuGlu 300
        ||||||||||||||||||||||||||||||||||:::::|||||||||||
    959 CGCTGGTAAAGAAGTACATGAAGCTGCAGCATGTCAACAAGAGGCTGGAG 1008

    301 ArgArgValIleAspTrpTyrGlnHisLeuGlnIleAsnLysLysMetTh 317
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1009 CGGCGAGTTATTGACTGGTACCAGCATCTTCAGATCAACAAGAAGATGAC 1058

    317 rAsnGluValAlaIleLeuGlnHisLeuProGluArgLeuArgAlaGluV 334
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1059 CAACGAGGTAGCCATCTTGCAGCACCTGCCTGAGCGGCTGCGGGCGGAGG 1108

    334 alAlaValSerValHisLeuSerThrLeuSerArgValGlnIlePheGln 350
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1109 TTGCTGTGTCCGTGCACCTGTCTACCCTGAGCCGAGTACAGATCTTCCAG 1158

    351 AsnCysGluAlaSerLeuLeuGluGluLeuValLeuLysLeuGlnProGl 367
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1159 AACTGTGAAGCCAGCCTGCTGGAAGAGCTGGTGCTGAAGCTACAGCCCCA 1208

    367 nThrTyrSerProGlyGluTyrValCysArgLysGlyAspIleGlyGlnG 384
        ||||||||||||||||||||||||||||||||||||||||||||||:::|
   1209 GACCTACTCGCCAGGCGAATATGTGTGCCGCAAAGGGGACATTGGCCGAG 1258

    384 luMetTyrIleIleArgGluGlyGlnLeuAlaValValAlaAspAspGly 400
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1259 AGATGTACATCATCCGTGAGGGCCAGCTGGCTGTGGTGGCCGATGATGGT 1308

    401 IleThrGlnTyrAlaValLeuGlyAlaGlyLeuTyrPheGlyGluIleSe 417
        :::||||||||||||||||||||||||||||||||||||||||||||||||
   1309 GTCACACAGTATGCTGTGCTTGGTGCAGGGCTCTACTTTGGGGAGATCAG 1358

    417 rIleIleAsnIleLysGlyAsnMetSerGlyAsnArgArgThrAlaAsnI 434
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1359 TATCATCAACATCAAAGGGAACATGTCTGGAAACCGACGAACAGCCAACA 1408

    434 leLysSerLeuGlyTyrSerAspLeuPheCysLeuSerLysGluAspLeu 450
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1409 TCAAGAGCCTAGGTTATTCAGACCTGTTCTGCCTCAGCAAGGAGGATCTG 1458

    451 ArgGluValLeuSerGluTyrProGlnAlaGlnThrIleMetGluGluLy 467
        ||||||||||||||||||||||||||||||||:::::|||||||||||||
   1459 CGGGAGGTACTGAGTGAGTACCCACAGGCCCAGGCGGTCATGGAGGAGAA 1508

    467 sGlyArgGluIleLeuLeuLysMetAsnLysLeuAspValAsnAlaGluA 484
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1509 GGGCCGAGAAATCTTGCTCAAAATGAATAAGTTGGATGTGAATGCTGAGG 1558

    484 laAlaGluIleAlaLeuGlnGluAlaThrGluSerArgLeuArgGlyLeu 500
        ||||||||||||||||||||||||||||||||||||||||||||:::|||||
   1559 CAGCTGAGATCGCCCTCCAGGAGGCCACAGAGTCTCGGCTCAAAGGCCTC 1608

    501 AspGlnGlnLeuAspAspLeuGlnThrLysPheAlaArgLeuLeuAlaGl 517
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1609 GACCAGCAGCTTGATGATCTGCAGACCAAGTTTGCTCGCCTACTGGCTGA 1658

    517 uLeuGluSerSerAlaLeuLysIleAlaTyrArgIleGluArgLeuGluT 534
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1659 GCTGGAGTCCAGTGCACTGAAGATAGCTTACCGCATCGAAAGGCTGGAGT 1708

    534 rpGlnThrArgGluTrpProMetProGluAspLeuAlaGluAlaAspAsp 550
        ||||||||||||||||||||||||||||||:::::|||||||||||||||
   1709 GGCAGACTCGAGAGTGGCCAATGCCAGAGGACATGGGTGAGGCTGATGAT 1758

    551 GluGlyGluProGluGluGlyThrSerLysAspGluGluGlyArgAlaSe 567
        |||:::||||||   ||||||||||||||||||   ||||||:::|||::
   1759 GAGGCTGAGCCTGGAGAAGGGACGTCCAAGGATGGAGAGGGAAAGGCTGG 1808

    567 rGlnGluGlyProProGlyProGlu 575
        :|||   ||||||   |||   |||
   1809 CCAGGCGGGACCCTCAGGCATAGAA 1833
```

seq_name: gb_htg:AC021935

seq_documentation_block:
LOCUS AC021935 162965 bp DNA linear HTG 03-MAR-2001

```
DEFINITION  Homo sapiens chromosome 11 clone RP11-645I8 map 11, WORKING DRAFT
            SEQUENCE, 5 unordered pieces.
ACCESSION   AC021935
VERSION     AC021935.3  GI:13184160
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 162965)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-645I8
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 162965)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 2, 2001 this sequence version replaced gi:7652890.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5277
                Center clone name: 645_I_8
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 40% of reads
                Sequencing vector: Plasmid; n/a; 60% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 161664 bases at least Q40
                Consensus quality: 161989 bases at least Q30
                Consensus quality: 162229 bases at least Q20
                Insert size: 158000; agarose-fp
                Insert size: 162565; sum-of-contigs
                Quality coverage: 12.2 in Q20 bases; agarose-fp
                Quality coverage: 11.9 in Q20.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     5772: contig of 5772 bp in length
            *     5773 5872: gap of      100 bp
            *     5873   114225: contig of 108353 bp in length
            *   114226 114325: gap of      100 bp
            *   114326   119496: contig of 5171 bp in length
            *   119497 119596: gap of      100 bp
            *   119597   161781: contig of 42185 bp in length
            *   161782 161881: gap of      100 bp
            *   161882   162965: contig of 1084 bp in length.
FEATURES             Location/Qualifiers
     source          1. .162965
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-645I8"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .5772
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    5873. .114225
                     /note="assembly_fragment"
     misc_feature    114326. .119496
                     /note="assembly_fragment"
     misc_feature    119597. .161781
                     /note="assembly_fragment"
     misc_feature    161882. .162965
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    54235 a  33886 c  32984 g  41460 t    400 others
ORIGIN


alignment_scores:
            Quality: 2266.00              Length:   1735
              Ratio:  3.955                 Gaps:      9
 Percent Similarity:  33.026   Percent Identity: 32.911

alignment_block:
 US-09-927-267-1 x AC021935   ..

 Align seg 1/1  to: AC021935  from: 1  to: 162965

      1 MetSerGlnAspThrLysValLysThrThrGluSerSerProProAlaPr 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
 149667 ATGAGCCAGGACACCAAAGTGAAGACAACAGAGTCCAGTCCCCCAGCCCC 149716

     17 oSerLysAla..................................... 20
        |||||||||
 149717 ATCCAAGGCCAGGTGAGAAGTCCTGGTCCCTTGTGTGGGATCTCTCCTCA 149766

     20 ................................................ 20

 149767 TTCCTCTTGGTGCCCCAGTCACAACTACAGCTTTGAATGCCTGGTGAATA 149816

     21 .....................Arg.LysLeuLeuProValLeuAspP 29
                              ::: |||||||||||||||||||||
 149817 AATGAAGCAAGACTTTCTTTCTTACAGGAAGTTGCTGCCTGTCCTGGACC 149866

     29 roSerGlyAspTyrTyrTyrTrpTrpLeuAsnThrMetValPheProVal 45
        |||||||||||||||||||||||||||||||||||||||||||||||||
 149867 CATCTGGGGATTACTACTACTGGTGGCTGAACACAATGGTCTTCCCAGTC 149916

     46 MetTyrAsnLeuIleIleLeuVal...................... 53
        |||||||||||||||||||||||
 149917 ATGTATAACCTCATCATCCTCGTGTGCAGGTATGGCAGCGGTGCTAAGGG 149966

     53 ................................................ 53

 149967 AGGGGCTGGAAGCCAAAAAGAGGACTAAAGAGAGGTCAAGGAGAAGGGCA 150016

     53 ................................................ 53

 150017 GACCCTTGGTGGGGCAGGAGGAGCAATTCCCATGGGAGGGCCTGAGGCAG 150066

     53 ................................................ 53

 150067 AGGGTTAAGGGCCCTGGGGAGACGCCTCGCACACAGAGGGTGCCCTTAAT 150116
```

```
    54 .....................CysArgAlaCysPheProAspLeuGlnH  63
                            ||||||||||||||||||||||||||||
150117 TCAATCATGCTTAACCCTGCCCTGCAGAGCCTGCTTCCCCGACTTGCAGC 150166

    63 isGlyTyrLeuValAlaTrpLeuValLeuAspTyrThrSerAspLeuLeu  79
       ||||||||||||||||||||||||||||||||||||||||||||||||||
150167 ACGGTTATCTGGTGGCCTGGTTGGTGCTGGACTACACGAGTGACCTGCTA 150216

    80 TyrLeuLeuAspMetValValArgPhe..His.....................  89
       |||||||||||||||||||||||||  |||
150217 TACCTACTAGACATGGTGGTGCGCTTCCACACAGGTCAGTGGGCTTCTAG 150266

    89 ..................................................... 89
       |||||||||||||||||||||||||||||||||||||||||||||||||||
150267 GAATGACCCTTTGTCCCACATTCCCTTCCTAAAGATAGCCACTTAAGAAG 150316

    89 ..................................................... 89
       |||||||||||||||||||||||||||||||||||||||||||||||||||
150317 TAACAAGAAAGGCACCCCCACCGTGGTAGCACCTTCGCGTGCCTCTATGC 150366

    89 ..................................................... 89
       |||||||||||||||||||||||||||||||||||||||||||||||||||
150367 CTGACAGCATCCCAGTGCTCACCCCGGAAAGCCGGGAGCAGAGTTATGCC 150416

    89 ..................................................... 89
       |||||||||||||||||||||||||||||||||||||||||||||||||||
150417 TGGCTCCACTCTGTCCTTCAGACAGTCTCCCTGGCCTGCCCTGGGCAGCT 150466

    89 ..................................................... 89
       |||||||||||||||||||||||||||||||||||||||||||||||||||
150467 CATGCTCAGCCCAAGCTTGACTACAGCAGGTCCGCTTCCTACCGGCTCCC 150516

    90 .....ThrGlyPheLeuGluGlnGlyIleLeuValValAspLysGlyArg 104
            ||||||||||||||||||||||||||||||||||||||||||||||
150517 TCTCCCCAGGATTCTTGGAACAGGGCATCCTGGTGGTGGACAAGGGTAGG 150566

   105 IleSerSerArgTyrValArgThrTrpSerPhePheLeuAspLeuAlaSe 121
       ||||||||||||||||||||||||||||||||||||||||||||||||||
150567 ATCTCGAGTCGCTACGTTCGCACCTGGAGTTTCTTCTTGGACCTGGCTTC 150616

   121 rLeuMetProThrAspValValTyrValArgLeuGlyProHisThrProT 138
       ||||||||||||||||||||||||||||||||||||||||||||||||||
150617 CCTGATGCCCACAGATGTGGTCTACGTGCGGCTGGGCCCGCACACACCCA 150666

   138 hrLeuArgLeuAsnArgPheLeuArgAlaProArgLeuPheGluAlaPhe 154
       ||||||||||||||||||||||||||||||||||||||||||||||||||
150667 CCCTGAGGCTGAACCGCTTTCTCCGCGCGCCCCGCCTCTTCGAGGCCTTC 150716

   155 AspArgThrGluThrArgThrAlaTyrProAsnAlaPheArgIleAlaLy 171
       ||||||||||||||||||||||||||||||||||||||||||||||||||
150717 GACCGCACAGAGACCCGCACAGCTTACCCAAATGCCTTTCGCATTGCCAA 150766

   171 sLeuMetLeuTyrIlePheValValIleHisTrpAsnSerCysLeuTyrP 188
       ||||||||||||||||||||||||||||||||||||||||||||||||||
150767 GCTGATGCTTTACATTTTTGTCGTCATCCATTGGAACAGCTGCCTATACT 150816

   188 heAlaLeuSerArgTyrLeuGlyPheGlyArgAspAlaTrpValTyrPro 204
       ||||||||||||||||||||||||||||||||||||||||||||||||||
150817 TTGCCCTATCCCGGTACCTGGGCTTCGGGCGTGACGCATGGGTGTACCCG 150866

   205 AspProAlaGlnProGlyPheGluArgLeuArgArgGlnTyrLeuTyrSe 221
       ||||||||||||||||||||||||||||||||||||||||||||||||||
150867 GACCCCGCGCAGCCTGGCTTTGAGCGCCTGCGGCGCCAGTACCTCTATAG 150916

   221 rPheTyrPheSerThrLeuIleLeuThrThrValGlyAspThrProProP 238
       ||||||||||||||||||||||||||||||||||||||||||||||||||
150917 CTTTTACTTCTCCACGCTGATACTGACTACAGTGGGCGATACACCGCCGC 150966

   238 roAlaArgGluGluGluTyrLeuPheMetValGlyAspPheLeuLeuAla 254
       ||||||||||||||||||||||||||||||||||||||||||||||||||
150967 CAGCCAGGGAAGAAGAGTACCTCTTCATGGTGGGCGACTTCCTGCTGGCC 151016

   255 ValMetGlyPheAlaThrIleMetGlySerMetSerSerValIleTyrAs 271
       ||||||||||||||||||||||||||||||||||||||||||||||||||
151017 GTCATGGGTTTCGCCACCATCATGGGTAGCATGAGCTCTGTCATCTACAA 151066

   271 nMetAsnThrAlaAspAlaAlaPheTyrProAspHisAlaLeuValLysL 288
       ||||||||||||||||||||||||||||||||||||||||||||||||||
151067 CATGAACACTGCAGATGCGGCTTTCTACCCAGATCATGCACTGGTGAAGA 151116

   288 ysTyrMetLysLeuGlnHisValAsnArgLysLeuGluArgArgValIle 304
       ||||||||||||||||||||||||||||||||||||||||||||||||||
151117 AGTACATGAAGCTGCAGCACGTCAACCGCAAGCTGGAGCGGCGAGTTATT 151166

   305 AspTrp..Tyr..................................... 307
       ||||||
151167 GACTGGTGAGAAGGCGGGGTTCCAGACCAGGACAGGGACCAGTGTAGGTG 151216

   307 ..................................................... 307
       |||||||||||||||||||||||||||||||||||||||||||||||||||
151217 ATGGAACCTGAGGGAGGTAACTGGGTCCTTAGTGCCTGGTGAGCCAGGCA 151266

   307 ..................................................... 307
       |||||||||||||||||||||||||||||||||||||||||||||||||||
151267 AGGCTGTCAAAATGTAGCATTCAGCCGTGGGTTTGCTGGCTGGGGCTTGG 151316

   307 ..................................................... 307
       |||||||||||||||||||||||||||||||||||||||||||||||||||
151317 AAAAAGGGAACTTCTTTCAACTGAGGGAATCAGGACTTGGGGGAGGGGTA 151366

   307 ..................................................... 307
       |||||||||||||||||||||||||||||||||||||||||||||||||||
151367 GGTAAGAGACTGATAGGGAGAGGAGCTCATACTCAAAAAAGGATAATATG 151416

   307 ..................................................... 307
       |||||||||||||||||||||||||||||||||||||||||||||||||||
151417 GAGACCAGGGAATGGGAAGTGCCACTGCCTGGTAGGGCTCAGAAGGCTCT 151466

   307 ..................................................... 307
       |||||||||||||||||||||||||||||||||||||||||||||||||||
151467 GGAAGGAGTGGGTGAAGAAGGGCAATCCCCCCTGAGAATGGTCAGCAACA 151516

   307 ..................................................... 307
       |||||||||||||||||||||||||||||||||||||||||||||||||||
151517 AGATCATTGGCTACCTATGGTGATTCATACTGTGGACTGAACAAAAGGAA 151566

   307 ..................................................... 307
       |||||||||||||||||||||||||||||||||||||||||||||||||||
151567 GTGTCACAAAGTTGGGGAAGTAGAGCAGAGGGTCCCCAGAGAGGAGGGGC 151616

   307 ..................................................... 307
       |||||||||||||||||||||||||||||||||||||||||||||||||||
151617 CCTTGGTGGCTGAGGTAGCTAAGGGTAGGGGTGGAACAAGACATCACTGA 151666

   307 ..................................................... 307
       |||||||||||||||||||||||||||||||||||||||||||||||||||
151667 TTGGTGGGTAGGGTAGAGCCTGATAGGAGGTGAAGGGTTATGTAAAGTGG 151716

   307 ..................................................... 307
       |||||||||||||||||||||||||||||||||||||||||||||||||||
151717 TAGAGGTGTCTATGCCCTGGAAAACAGGTAATCCAACTGTCAGGTACTCC 151766

   307 ..................................................... 307
       |||||||||||||||||||||||||||||||||||||||||||||||||||
151767 CATGACCCCTGTTAGATCTCAGTATGGTGGGATTCTTGGCTGGAGCTGAG 151816

   307 ..................................................... 307
       |||||||||||||||||||||||||||||||||||||||||||||||||||
151817 CTGAGCCCTGAAGGAAGGGATGGAGGCAGGCACATAAAGGGCTGGTAAGG 151866

   308 ..........................GlnHisLeuGlnIleAsnLysL 315
                                ||||||||||||||||||||||
151867 AAATGGCACTGTCCTTACTCTCAGGTATCAGCACCTGCAGATCAACAAGA 151916

   315 ysMetThrAsnGluValAlaIleLeuGlnHisLeuProGluArgLeuArg 331
```

```
151917 AGATGACCAACGAGGTAGCCATCTTACAGCACTTGCCTGAGCGGCTGCGG 151966

   332 AlaGluValAlaValSerValHisLeuSerThrLeuSerArgValGlnIl 348
151967 GCAGAAGTGGCTGTGTCTGTGCACCTGTCCACTCTGAGCCGGGTGCAGAT 152016

   348 ePheGlnAsnCysGluAlaSerLeuLeuGluGluLeuValLeuLysLeuG 365
152017 CTTTCAGAACTGTGAGGCCAGCCTGCTGGAGGAGCTGGTGCTGAAGCTGC 152066

   365 lnProGlnThrTyrSerProGlyGluTyrValCysArgLysGlyAspIle 381
152067 AGCCCCAGACCTACTCACCAGGTGAATATGTATGCCGCAAAGGAGACATT 152116

   382 GlyGlnGluMetTyrIleIleArgGluGlyGlnLeuAlaValValAlaAs 398
152117 GGCCAAGAGATGTACATCATCCGAGAGGGTCAACTGGCCGTGGTGGCAGA 152166

   398 pAspGlyIleThrGlnTyrAlaValLeuGlyAlaGlyLeuTyrPheGlyG 415
152167 TGATGGTATCACACAGTATGCTGTGCTCGGTGCAGGGCTCTACTTTGGGG 152216

   415 luIleSerIleIleAsnIle..Lys..... 422
152217 AGATCAGCATCATCAACATCAAAGGTGGGTATCCCAGTATTTGTTCCAGG 152266

   422 ..... 422

152267 GACAAGGATGGGTGGGGTAGGGGGGAACAGCAGAGCCCAGTGCTGGGACC 152316

   422 ..... 422

152317 AGATAGTACTTCAGGCCTAAACTTCTGATTGAGGAAACCTGGCCCTTCTC 152366

   422 ..... 422

152367 TGAGTCACTAGATGGCTCAGGAGAAAAACAACACATAGAGGATCCATTCC 152416

   422 ..... 422

152417 CTGAGAAGAAGCTGAGCCAAGGCCAGTGAACAGGGTGGGTTCGTGGAAGA 152466

   422 ..... 422

152467 GAGAGTAGACTTGCTCTGTGGATATTCACCTGAAGCACCAATAGACTAGT 152516

   422 ..... 422

152517 GGACAGTGATAGAGATAATAATAATAATAATAGCTAATATTTATTACCTG 152566

   422 ..... 422

152567 CTTACTGCATCCCCTTTTCTAAGTGCTTCACATGATATTAACTCATTACA 152616

   422 ..... 422

152617 TTCCCGCAGCAGTTCTATAAGATAAGTACTATTCTTATCCCAATTTTAAA 152666

   422 ..... 422

152667 GATAAGGAAATCAAACTCAAAGAAGTTGGGTAATTTGCCCAACGTCGCAC 152716

   422 ..... 422

152717 AGGTATACACATGTGGGATGAGTGGAAAAAACAACCTTTTTACAGCCAAG 152766

   422 ..... 422

152767 TGGAAAGAGAAAGGTAGAAGAGAAGAAGTTCCCATTTACTGACCATCACC 152816

   422 ..... 422

152817 TCTAGGCCAAGCCAGGTACTACCTACTGGCTTGCTTTTTCCTCTCATCAG 152866

   422 ..... 422

152867 CCCTGTGCAGTAGGAATTGTTATGCCTAATCACAGGTATTAGGCATGAGG 152916

   422 ..... 422

152917 TCGTGCAAGCAGGAAATCCTGCAAATGGGGTAAGCACACTGTCATGGACC 152966

   422 ..... 422

152967 GAGAAGCTGGATGTGCTCTGGCTCTCTCAGCATGAGGATTCTACCCTCAG 153016

   422 ..... 422

153017 GGATTCTGAAAGCCTCACATCTGTCCTCCAATGATCCCCTCAAGCTCCAC 153066

   422 ..... 422

153067 ATCCTGTATGTGGTAGGATGCTGTCTGTTTGACTTTGGACAAGTCATTTC 153116

   422 ..... 422

153117 CCTATTTAGGCCTCAGTTTAATCACCTATCCAATGAGGACAATAATAAAA 153166

   422 ..... 422

153167 ACCCAATGAGTTATTTGTGAGAATAAATGAGCTATGTGAGTGAGGCATCC 153216

   422 ..... 422

153217 AGCGCAGTGCTTGGCATACAGCAAGTGTTCAACAAATAGTAGCTCCCTTT 153266

   422 ..... 422

153267 CCCTTGTCCAAGGTCCTTCTAGCTCTAAGACTTTGCAGCATTCAGCCTCA 153316

   422 ..... 422

153317 CCTATCTCCCCACCCATTCTCCACCAACATCTTTCTAAACTGCAAAGGAT 153366

   422 ..... 422

153367 ATCATACAGCTCCCTGCTTAAAATACTTCAATAGTTTCCTACTGCCCTCA 153416

   422 ..... 422

153417 GCATAAGGCCTGAACTTCATGTCATGGCTTATGAGACCCAGTGTGACCTA 153466

   422 ..... 422

153467 GATCTCCCCTCTCTCCACCCTCCCCACACTCTGCGCTTCTCGCACTCTGA 153516

   422 ..... 422

153517 ACTGCTTACTATTTTCTGCATCCAACTGACTCTTTTCTACCCCCTTCCCT 153566

   422 ..... 422

153567 CTGTTCCTGCTCCTTCCTCTGGCTAGATTGCCCTATCCCCACTTGTCCCC 153616

   422 ..... 422

153617 CTCCTCCCCCTAGTTACCTCCTACTCAGTTCAGGTGTATGAGACTGTTCT 153666

   422 ..... 422

153667 TGCATTTCTATAAAGAAATATGTGAGACTGGATAATTTATAAAGAAAAGA 153716

   422 ..... 422

153717 GGTTTAATTGGCTCACAGTTCTTCAGGATTTACGGGAAGCATGGTGCTGG 153766
```

```
   422 ..................................................  422

153767 CATTTGCTCAGCTTCTAGGGAGGCTTCAAGAAGCTTATAATTGTGGCAGA 153816

   422 ..................................................  422

153817 AGGCAAAGGGGGAGCAGGCATGTCACACGGTGAAAGCAGGAGCAAGGGTT 153866

   422 ..................................................  422

153867 GGGGGAGGTGCCACACTTTCAAACAACCAAAACAACCAGCTCTCGACTCA 153916

   422 ..................................................  422

153917 CTAACTCAAAGACAGCATCAAGCCATGAAGGATTCGCCCCCATAATCCAA 153966

   422 ..................................................  422

153967 TCACCTCCCACCAGGTCCCACCCCCAACATTGGGGATTACATTTCACATG 154016

   422 ..................................................  422

154017 AGATTTGGGCAGAGACAAATATTCAAATTATATCTTAGGACATCCCTTCC 154066

   422 ..................................................  422

154067 TCCAGACTTCCCTAAATTCCCTGCCTTACGGTTTGGTAGGGGCTCTTTGC 154116

   422 ..................................................  422

154117 CTACCTTTCCACAGCACCTGATGAACATGTCTTCACTGCACCAGCCATAC 154166

   422 ..................................................  422

154167 TGTTATATAACTATTCCAATATATGTGTATCTCCTCTAGACTGTGAATTA 154216

   422 ..................................................  422

154217 TTTGAGGCAGGTCACTGATACCTACCCAGCATGGAGCCTGGTCCCCAGTA 154266

   422 ..................................................  422

154267 TGTTACTAAATGAAAGAATTAATGAGGCAGAAGGAGAGGCTCAGAAGCAC 154316

   422 ..................................................  422

154317 AAATATGGAGGTGAAGGTCCTGGTTCAGGAGTGAAATGCCACCTCCTCAC 154366

   423 .....................................GlyAsnMet  425
                                                 |||||||||
154367 CCTCCTACTAACTGTCCTCCCATCTCTGCCCATGAGCCACAGGGAACATG 154416

   426 SerGlyAsnArgArgThrAlaAsnIleLysSerLeuGlyTyrSerAspLe 442
       ||||||||||||||||||||||||||||||||||||||||||||||||||
154417 TCTGGGAACCGCCGCACAGCCAACATCAAGAGCCTAGGTTATTCAGACCT 154466

   442 uPheCysLeuSerLysGluAspLeuArgGluValLeuSerGluTyrProG 459
       ||||||||||||||||||||||||||||||||||||||||||||||||||
154467 ATTCTGCCTGAGCAAGGAGGACCTGCGGGAGGTGCTGAGCGAGTATCCAC 154516

   459 lnAlaGlnThrIleMetGluGluLysGlyArgGluIleLeuLeuLysMet 475
       ||||||||||||||||||||||||||||||||||||||||||||||||||
154517 AAGCACAGACCATCATGGAGGAGAAAGGACGTGAGATCCTGCTGAAAATG 154566

   476 AsnLysLeuAspValAsnAlaGluAlaAlaGluIleAlaLeuGlnGluAl 492
       ||||||||||||||||||||||||||||||||||||||||||||||||||
154567 AACAAGTTGGACGTGAATGCTGAGGCAGCTGAGATCGCCCTGCAGGAGGC 154616

   492 aThrGluSerArgLeuArgGlyLeuAspGlnGlnLeuAspAspLeuGlnT 509
       ||||||||||||||||||||||||||||||||||||||||||||||||||
154617 CACAGAGTCCCGGCTACGAGGCCTAGACCAGCAGCTGGATGATCTACAGA 154666

   509 hrLysPheAlaArgLeuLeuAlaGluLeuGluSerSerAlaLeuLysIle 525
       ||||||||||||||||||||||||||||||||||||||||||||||||||
154667 CCAAGTTTGCTCGCCTCCTGGCTGAGCTGGAGTCCAGCGCACTTAAGATT 154716

   526 AlaTyrArgIleGluArgLeuGluTrpGlnThrArgGluTrpProMetPr 542
       ||||||||||||||||||||||||||||||||||||||||||||||||||
154717 GCTTACCGCATTGAACGGCTGGAGTGGCAGACTCGAGAGTGGCCAATGCC 154766

   542 oGluAspLeuAlaGluAlaAspAspGluGlyGluProGluGluGlyThrS 559
       ||||||||||||||||||||||||||||||||||||||||||||||||||
154767 CGAGGACCTGGCTGAGGCTGATGACGAGGGTGAGCCTGAGGAGGGAACTT 154816

   559 erLysAspGluGluGlyArgAlaSerGlnGluGlyProProGlyProGlu 575
       ||||||||||||||||||||||||||||||||||||||||||||||||||
154817 CCAAAGATGAAGAGGGCAGGGCCAGCCAGGAGGGACCCCCAGGTCCAGAG 154866

seq_name: gb_htg:AC022762

seq_documentation_block:
LOCUS       AC022762              186609 bp    DNA     linear   HTG 09-FEB-2002
DEFINITION  Homo sapiens chromosome 11 clone RP11-290F24 map 11, *** SEQUENCING
            IN PROGRESS ***, 1 ordered pieces.
ACCESSION   AC022762
VERSION     AC022762.6  GI:18642872
KEYWORDS    HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 186609)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-290F24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 186609)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 9, 2002 this sequence version replaced gi:18483561.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5842
                Center clone name: 290_F_24
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
```

```
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
          *        1   186609: contig of 186609 bp in length.
FEATURES             Location/Qualifiers
     source          1. .186609
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-290F24"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    60812 a  39184 c  37829 g  48784 t
ORIGIN


alignment_scores:
           Quality: 2266.00            Length:   1735
             Ratio:  3.955              Gaps:     9
 Percent Similarity:  33.026   Percent Identity: 32.911

alignment_block:
 US-09-927-267-1 x AC022762   ..

 Align seg 1/1  to: AC022762  from: 1  to: 186609

       1 MetSerGlnAspThrLysValLysThrThrGluSerSerProProAlaPr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  155446 ATGAGCCAGGACACCAAAGTGAAGACAACAGAGTCCAGTCCCCCAGCCCC 155495

      17 oSerLysAla..................................... 20
         |||||||||
  155496 ATCCAAGGCCAGGTGAGAAGTCCTGGTCCCTTGTGTGGGATCTCTCCTCA 155545

      20 ..................................................  20

  155546 TTCCTCTTGGTGCCCCAGTCACAACTACAGCTTTGAATGCCTGGTGAATA 155595

      21 .....................Arg.LysLeuLeuProValLeuAspP 29
                               ::: |||||||||||||||||||||
  155596 AATGAAGCAAGACTTTCTTTCTTACAGGAAGTTGCTGCCTGTCCTGGACC 155645

      29 roSerGlyAspTyrTyrTyrTrpTrpLeuAsnThrMetValPheProVal 45
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  155646 CATCTGGGGATTACTACTACTGGTGGCTGAACACAATGGTCTTCCCAGTC 155695

      46 MetTyrAsnLeuIleIleLeuVal...................... 53
         |||||||||||||||||||||||
  155696 ATGTATAACCTCATCATCCTCGTGTGCAGGTATGGCAGCGGTGCTAAGGG 155745

      53 ..................................................  53

  155746 AGGGGCTGGAAGCCAAAAAGAGGACTAAAGAGAGGTCAAGGAGAAGGGCA 155795

      53 ..................................................  53

  155796 GACCCTTGGTGGGGCAGGAGGAGCAATTCCCATGGGAGGGCCTGAGGCAG 155845

      53 ..................................................  53

  155846 AGGGTTAAGGGCCCTGGGGAGACGCCTCGCACACAGAGGGTGCCCTTAAT 155895

      54 .....................CysArgAlaCysPheProAspLeuGlnH 63
                              ||||||||||||||||||||||||||||
  155896 TCAATCATGCTTAACCCTGCCCTGCAGAGCCTGCTTCCCCGACTTGCAGC 155945

      63 isGlyTyrLeuValAlaTrpLeuValLeuAspTyrThrSerAspLeuLeu 79
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  155946 ACGGTTATCTGGTGGCCTGGTTGGTGCTGGACTACACGAGTGACCTGCTA 155995

      80 TyrLeuLeuAspMetValValArgPhe..His..................  89
         |||||||||||||||||||||||||||  |||
  155996 TACCTACTAGACATGGTGGTGCGCTTCCACACAGGTCAGTGGGCTTCTAG 156045

      89 ..................................................  89

  156046 GAATGACCCTTTGTCCCACATTCCCTTCCTAAAGATAGCCACTTAAGAAG 156095

      89 ..................................................  89

  156096 TAACAAGAAAGGCACCCCCACCGTGGTAGCACCTTCGCGTGCCTCTATGC 156145

      89 ..................................................  89

  156146 CTGACAGCATCCCAGTGCTCACCCCGGAAAGCCGGGAGCAGAGTTATGCC 156195

      89 ..................................................  89

  156196 TGGCTCCACTCTGTCCTTCAGACAGTCTCCCTGGCCTGCCCTGGGCAGCT 156245

      89 ..................................................  89

  156246 CATGCTCAGCCCAAGCTTGACTACAGCAGGTCCGCTTCCTACCGGCTCCC 156295

      90 .....ThrGlyPheLeuGluGlnGlyIleLeuValValAspLysGlyArg 104
               ||||||||||||||||||||||||||||||||||||||||||||||||||
  156296 TCTCCCCAGGATTCTTGGAACAGGGCATCCTGGTGGTGGACAAGGGTAGG 156345

     105 IleSerSerArgTyrValArgThrTrpSerPhePheLeuAspLeuAlaSe 121
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  156346 ATCTCGAGTCGCTACGTTCGCACCTGGAGTTTCTTCTTGGACCTGGCTTC 156395

     121 rLeuMetProThrAspValValTyrValArgLeuGlyProHisThrProT 138
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  156396 CCTGATGCCCACAGATGTGGTCTACGTGCGGCTGGGCCCGCACACACCCA 156445

     138 hrLeuArgLeuAsnArgPheLeuArgAlaProArgLeuPheGluAlaPhe 154
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  156446 CCCTGAGGCTGAACCGCTTTCTCCGCGCGCCCCGCCTCTTCGAGGCCTTC 156495

     155 AspArgThrGluThrArgThrAlaTyrProAsnAlaPheArgIleAlaLy 171
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  156496 GACCGCACAGAGACCCGCACAGCTTACCCAAATGCCTTTCGCATTGCCAA 156545

     171 sLeuMetLeuTyrIlePheValValIleHisTrpAsnSerCysLeuTyrP 188
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  156546 GCTGATGCTTTACATTTTTGTCGTCATCCATTGGAACAGCTGCCTATACT 156595

     188 heAlaLeuSerArgTyrLeuGlyPheGlyArgAspAlaTrpValTyrPro 204
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  156596 TTGCCCTATCCCGGTACCTGGGCTTCGGGCGTGACGCATGGGTGTACCCG 156645

     205 AspProAlaGlnProGlyPheGluArgLeuArgArgGlnTyrLeuTyrSe 221
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  156646 GACCCCGCGCAGCCTGGCTTTGAGCGCCTGCGGCGCCAGTACCTCTATAG 156695

     221 rPheTyrPheSerThrLeuIleLeuThrThrValGlyAspThrProProP 238
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  156696 CTTTTACTTCTCCACGCTGATACTGACTACAGTGGGCGATACACCGCCGC 156745

     238 roAlaArgGluGluGluTyrLeuPheMetValGlyAspPheLeuLeuAla 254
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  156746 CAGCCAGGGAAGAAGAGTACCTCTTCATGGTGGGCGACTTCCTGCTGGCC 156795

     255 ValMetGlyPheAlaThrIleMetGlySerMetSerSerValIleTyrAs 271
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  156796 GTCATGGGTTTCGCCACCATCATGGGTAGCATGAGCTCTGTCATCTACAA 156845

     271 nMetAsnThrAlaAspAlaAlaPheTyrProAspHisAlaLeuValLysL 288
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  156846 CATGAACACTGCAGATGCGGCTTTCTACCCAGATCATGCACTGGTGAAGA 156895

     288 ysTyrMetLysLeuGlnHisValAsnArgLysLeuGluArgArgValIle 304
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  156896 AGTACATGAAGCTGCAGCACGTCAACCGCAAGCTGGAGCGGCGAGTTATT 156945
```

```
   305 AspTrp..Tyr..................................... 307
       ||||||
156946 GACTGGTGAGAAGGCGGGGTTCCAGACCAGGACAGGGACCAGTGTAGGTG 156995

   307 ..................................................... 307

156996 ATGGAACCTGAGGGAGGTAACTGGGTCCTTAGTGCCTGGTGAGCCAGGCA 157045

   307 ..................................................... 307

157046 AGGCTGTCAAAATGTAGCATTCAGCCGTGGGTTTGCTGGCTGGGGCTTGG 157095

   307 ..................................................... 307

157096 AAAAAGGGAACTTCTTTCAACTGAGGGAATCAGGACTTGGGGGAGGGGTA 157145

   307 ..................................................... 307

157146 GGTAAGAGACTGATAGGGAGAGGAGCTCATACTCAAAAAAGGATAATATG 157195

   307 ..................................................... 307

157196 GAGACCAGGGAATGGGAAGTGCCACTGCCTGGTAGGGCTCAGAAGGCTCT 157245

   307 ..................................................... 307

157246 GGAAGGAGTGGGTGAAGAAGGGCAATCCCCCCTGAGAATGGTCAGCAACA 157295

   307 ..................................................... 307

157296 AGATCATTGGCTACCTATGGTGATTCATACTGTGGACTGAACAAAAGGAA 157345

   307 ..................................................... 307

157346 GTGTCACAAAGTTGGGGAAGTAGAGCAGAGGGTCCCCAGAGAGGAGGGGC 157395

   307 ..................................................... 307

157396 CCTTGGTGGCTGAGGTAGCTAAGGGTAGGGGTGGAACAAGACATCACTGA 157445

   307 ..................................................... 307

157446 TTGGTGGGTAGGGTAGAGCCTGATAGGAGGTGAAGGGTTATGTAAAGTGG 157495

   307 ..................................................... 307

157496 TAGAGGTGTCTATGCCCTGGAAAACAGGTAATCCAACTGTCAGGTACTCC 157545

   307 ..................................................... 307

157546 CATGACCCCTGTTAGATCTCAGTATGGTGGGATTCTTGGCTGGAGCTGAG 157595

   307 ..................................................... 307

157596 CTGAGCCCTGAAGGAAGGGATGGAGGCAGGCACATAAAGGGCTGGTAAGG 157645

   308 ..........................GlnHisLeuGlnIleAsnLysL 315
                                ||||||||||||||||||||||
157646 AAATGGCACTGTCCTTACTCTCAGGTATCAGCACCTGCAGATCAACAAGA 157695

   315 ysMetThrAsnGluValAlaIleLeuGlnHisLeuProGluArgLeuArg 331
       ||||||||||||||||||||||||||||||||||||||||||||||||||
157696 AGATGACCAACGAGGTAGCCATCTTACAGCACTTGCCTGAGCGGCTGCGG 157745

   332 AlaGluValAlaValSerValHisLeuSerThrLeuSerArgValGlnIl 348
       ||||||||||||||||||||||||||||||||||||||||||||||||||
157746 GCAGAAGTGGCTGTGTCTGTGCACCTGTCCACTCTGAGCCGGGTGCAGAT 157795

   348 ePheGlnAsnCysGluAlaSerLeuLeuGluGluLeuValLeuLysLeuG 365
       ||||||||||||||||||||||||||||||||||||||||||||||||||
157796 CTTTCAGAACTGTGAGGCCAGCCTGCTGGAGGAGCTGGTGCTGAAGCTGC 157845

   365 lnProGlnThrTyrSerProGlyGluTyrValCysArgLysGlyAspIle 381
```

```
       ||||||||||||||||||||||||||||||||||||||||||||||||||
157846 AGCCCCAGACCTACTCACCAGGTGAATATGTATGCCGCAAAGGAGACATT 157895

   382 GlyGlnGluMetTyrIleIleArgGluGlyGlnLeuAlaValValAlaAs 398
       ||||||||||||||||||||||||||||||||||||||||||||||||||
157896 GGCCAAGAGATGTACATCATCCGAGAGGGTCAACTGGCCGTGGTGGCAGA 157945

   398 pAspGlyIleThrGlnTyrAlaValLeuGlyAlaGlyLeuTyrPheGlyG 415
       ||||||||||||||||||||||||||||||||||||||||||||||||||
157946 TGATGGTATCACACAGTATGCTGTGCTCGGTGCAGGGCTCTACTTTGGGG 157995

   415 luIleSerIleIleAsnIle..Lys..................... 422
       ||||||||||||||||||  :::
157996 AGATCAGCATCATCAACATCAAAGGTGGGTATCCCAGTATTTGTTCCAGG 158045

   422 ..................................................... 422

158046 GACAAGGATGGGTGGGGTAGGGGGGAACAGCAGAGCCCAGTGCTGGGACC 158095

   422 ..................................................... 422

158096 AGATAGTACTTCAGGCCTAAACTTCTGATTGAGGAAACCTGGCCCTTCTC 158145

   422 ..................................................... 422

158146 TGAGTCACTAGATGGCTCAGGAGAAAAACAACACATAGAGGATCCATTCC 158195

   422 ..................................................... 422

158196 CTGAGAAGAAGCTGAGCCAAGGCCAGTGAACAGGGTGGGTTCGTGGAAGA 158245

   422 ..................................................... 422

158246 GAGAGTAGACTTGCTCTGTGGATATTCACCTGAAGCACCAATAGACTAGT 158295

   422 ..................................................... 422

158296 GGACAGTGATAGAGATAATAATAATAATAATAGCTAATATTTATTACCTG 158345

   422 ..................................................... 422

158346 CTTACTGCATCCCCTTTTCTAAGTGCTTCACATGATATTAACTCATTACA 158395

   422 ..................................................... 422

158396 TTCCCGCAGCAGTTCTATAAGATAAGTACTATTCTTATCCCAATTTTAAA 158445

   422 ..................................................... 422

158446 GATAAGGAAATCAAACTCAAAGAAGTTGGGTAATTTGCCCAACGTCGCAC 158495

   422 ..................................................... 422

158496 AGGTATACACATGTGGGATGAGTGGAAAAAACAACCTTTTTACAGCCAAG 158545

   422 ..................................................... 422

158546 TGGAAAGAGAAAGGTAGAAGAGAAGAAGTTCCCATTTACTGACCATCACC 158595

   422 ..................................................... 422

158596 TCTAGGCCAAGCCAGGTACTACCTACTGGCTTGCTTTTTCCTCTCATCAG 158645

   422 ..................................................... 422

158646 CCCTGTGCAGTAGGAATTGTTATGCCTAATCACAGGTATTAGGCATGAGG 158695

   422 ..................................................... 422

158696 TCGTGCAAGCAGGAAATCCTGCAAATGGGGTAAGCACACTGTCATGGACC 158745

   422 ..................................................... 422
```

```
158746 GAGAAGCTGGATGTGCTCTGGCTCTCTCAGCATGAGGATTCTACCCTCAG 158795
   422 ..................................................  422
158796 GGATTCTGAAAGCCTCACATCTGTCCTCCAATGATCCCCTCAAGCTCCAC 158845
   422 ..................................................  422
158846 ATCCTGTATGTGGTAGGATGCTGTCTGTTTGACTTTGGACAAGTCATTTC 158895
   422 ..................................................  422
158896 CCTATTTAGGCCTCAGTTTAATCACCTATCCAATGAGGACAATAATAAAA 158945
   422 ..................................................  422
158946 ACCCAATGAGTTATTTGTGAGAATAAATGAGCTATGTGAGTGAGGCATCC 158995
   422 ..................................................  422
158996 AGCGCAGTGCTTGGCATACAGCAAGTGTTCAACAAATAGTAGCTCCCTTT 159045
   422 ..................................................  422
159046 CCCTTGTCCAAGGTCCTTCTAGCTCTAAGACTTTGCAGCATTCAGCCTCA 159095
   422 ..................................................  422
159096 CCTATCTCCCCACCCATTCTCCACCAACATCTTTCTAAACTGCAAAGGAT 159145
   422 ..................................................  422
159146 ATCATACAGCTCCCTGCTTAAAATACTTCAATAGTTTCCTACTGCCCTCA 159195
   422 ..................................................  422
159196 GCATAAGGCCTGAACTTCATGTCATGGCTTATGAGACCCAGTGTGACCTA 159245
   422 ..................................................  422
159246 GATCTCCCCTCTCTCCACCCTCCCCACACTCTGCGCTTCTCGCACTCTGA 159295
   422 ..................................................  422
159296 ACTGCTTACTATTTTCTGCATCCAACTGACTCTTTTCTACCCCCTTCCCT 159345
   422 ..................................................  422
159346 CTGTTCCTGCTCCTTCCTCTGGCTAGATTGCCCTATCCCCACTTGTCCCC 159395
   422 ..................................................  422
159396 CTCCTCCCCCTAGTTACCTCCTACTCAGTTCAGGTGTATGAGACTGTTCT 159445
   422 ..................................................  422
159446 TGCATTTCTATAAAGAAATATGTGAGACTGGATAATTTATAAAGAAAAGA 159495
   422 ..................................................  422
159496 GGTTTAATTGGCTCACAGTTCTTCAGGATTTACGGGAAGCATGGTGCTGG 159545
   422 ..................................................  422
159546 CATTTGCTCAGCTTCTAGGGAGGCTTCAAGAAGCTTATAATTGTGGCAGA 159595
   422 ..................................................  422
159596 AGGCAAAGGGGGAGCAGGCATGTCACACGGTGAAAGCAGGAGCAAGGGTT 159645
   422 ..................................................  422
159646 GGGGGAGGTGCCACACTTTCAAACAACCAAAACAACCAGCTCTCGACTCA 159695
```

```
   422 ..................................................  422
159696 CTAACTCAAAGACAGCATCAAGCCATGAAGGATTCGCCCCCATAATCCAA 159745
   422 ..................................................  422
159746 TCACCTCCCACCAGGTCCCACCCCCAACATTGGGGATTACATTTCACATG 159795
   422 ..................................................  422
159796 AGATTTGGGCAGAGACAAATATTCAAATTATATCTTAGGACATCCCTTCC 159845
   422 ..................................................  422
159846 TCCAGACTTCCCTAAATTCCCTGCCTTACGGTTTGGTAGGGGCTCTTTGC 159895
   422 ..................................................  422
159896 CTACCTTTCCACAGCACCTGATGAACATGTCTTCACTGCACCAGCCATAC 159945
   422 ..................................................  422
159946 TGTTATATAACTATTCCAATATATGTGTATCTCCTCTAGACTGTGAATTA 159995
   422 ..................................................  422
159996 TTTGAGGCAGGTCACTGATACCTACCCAGCATGGAGCCTGGTCCCCAGTA 160045
   422 ..................................................  422
160046 TGTTACTAAATGAAAGAATTAATGAGGCAGAAGGAGAGGCTCAGAAGCAC 160095
   422 ..................................................  422
160096 AAATATGGAGGTGAAGGTCCTGGTTCAGGAGTGAAATGCCACCTCCTCAC 160145
   423 .....................................GlyAsnMet  425
                                                |||||||||
160146 CCTCCTACTAACTGTCCTCCCATCTCTGCCCATGAGCCACAGGGAACATG 160195
   426 SerGlyAsnArgArgThrAlaAsnIleLysSerLeuGlyTyrSerAspLe 442
       ||||||||||||||||||||||||||||||||||||||||||||||||||
160196 TCTGGGAACCGCCGCACAGCCAACATCAAGAGCCTAGGTTATTCAGACCT 160245
   442 uPheCysLeuSerLysGluAspLeuArgGluValLeuSerGluTyrProG 459
       ||||||||||||||||||||||||||||||||||||||||||||||||||
160246 ATTCTGCCTGAGCAAGGAGGACCTGCGGGAGGTGCTGAGCGAGTATCCAC 160295
   459 lnAlaGlnThrIleMetGluGluLysGlyArgGluIleLeuLeuLysMet 475
       ||||||||||||||||||||||||||||||||||||||||||||||||||
160296 AAGCACAGACCATCATGGAGGAGAAAGGACGTGAGATCCTGCTGAAAATG 160345
   476 AsnLysLeuAspValAsnAlaGluAlaAlaGluIleAlaLeuGlnGluAl 492
       ||||||||||||||||||||||||||||||||||||||||||||||||||
160346 AACAAGTTGGACGTGAATGCTGAGGCAGCTGAGATCGCCCTGCAGGAGGC 160395
   492 aThrGluSerArgLeuArgGlyLeuAspGlnGlnLeuAspAspLeuGlnT 509
       ||||||||||||||||||||||||||||||||||||||||||||||||||
160396 CACAGAGTCCCGGCTACGAGGCCTAGACCAGCAGCTGGATGATCTACAGA 160445
   509 hrLysPheAlaArgLeuLeuAlaGluLeuGluSerSerAlaLeuLysIle 525
       ||||||||||||||||||||||||||||||||||||||||||||||||||
160446 CCAAGTTTGCTCGCCTCCTGGCTGAGCTGGAGTCCAGCGCACTTAAGATT 160495
   526 AlaTyrArgIleGluArgLeuGluTrpGlnThrArgGluTrpProMetPr 542
       ||||||||||||||||||||||||||||||||||||||||||||||||||
160496 GCTTACCGCATTGAACGGCTGGAGTGGCAGACTCGAGAGTGGCCAATGCC 160545
   542 oGluAspLeuAlaGluAlaAspAspGluGlyGluProGluGluGlyThrS 559
       ||||||||||||||||||||||||||||||||||||||||||||||||||
160546 CGAGGACCTGGCTGAGGCTGATGACGAGGGTGAGCCTGAGGAGGGAACTT 160595
```

```
     559 erLysAspGluGluGlyArgAlaSerGlnGluGlyProProGlyProGlu 575
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  160596 CCAAAGATGAAGAGGGCAGGGCCAGCCAGGAGGGACCCCCAGGTCCAGAG 160645

seq_name: gb_htg:AC036216

seq_documentation_block:
LOCUS       AC036216              164434 bp    DNA     linear   HTG 04-MAY-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-658A5 map 11, WORKING DRAFT
            SEQUENCE, 26 unordered pieces.
ACCESSION   AC036216
VERSION     AC036216.2  GI:7705203
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 164434)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-658A5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 164434)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 4, 2000 this sequence version replaced gi:7523886.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L9404
                Center clone name: 658_A_5
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 149016 bases at least Q40
                Consensus quality: 157301 bases at least Q30
                Consensus quality: 160298 bases at least Q20
                Insert size: 157000; agarose-fp
                Insert size: 161934; sum-of-contigs
                Quality coverage: 3.7 in Q20 bases; agarose-fp
                Quality coverage: 3.6 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1416: contig of 1416 bp in length
            *     1417  1516: gap of      100 bp
            *     1517     3580: contig of 2064 bp in length
            *     3581  3680: gap of      100 bp
            *     3681     4383: contig of 703 bp in length
            *     4384  4483: gap of      100 bp
            *     4484     5844: contig of 1361 bp in length
            *     5845  5944: gap of      100 bp
            *     5945     7574: contig of 1630 bp in length
            *     7575  7674: gap of      100 bp
            *     7675     9988: contig of 2314 bp in length
            *     9989 10088: gap of      100 bp
            *    10089    13426: contig of 3338 bp in length
            *    13427 13526: gap of      100 bp
            *    13527    15573: contig of 2047 bp in length
            *    15574 15673: gap of      100 bp
            *    15674    18974: contig of 3301 bp in length
            *    18975 19074: gap of      100 bp
            *    19075    21809: contig of 2735 bp in length
            *    21810 21909: gap of      100 bp
            *    21910    25085: contig of 3176 bp in length
            *    25086 25185: gap of      100 bp
            *    25186    29425: contig of 4240 bp in length
            *    29426 29525: gap of      100 bp
            *    29526    34133: contig of 4608 bp in length
            *    34134 34233: gap of      100 bp
            *    34234    40649: contig of 6416 bp in length
            *    40650 40749: gap of      100 bp
            *    40750    45891: contig of 5142 bp in length
            *    45892 45991: gap of      100 bp
            *    45992    53180: contig of 7189 bp in length
            *    53181 53280: gap of      100 bp
            *    53281    62817: contig of 9537 bp in length
            *    62818 62917: gap of      100 bp
            *    62918    70247: contig of 7330 bp in length
            *    70248 70347: gap of      100 bp
            *    70348    78754: contig of 8407 bp in length
            *    78755 78854: gap of      100 bp
            *    78855    86840: contig of 7986 bp in length
            *    86841 86940: gap of      100 bp
            *    86941    98610: contig of 11670 bp in length
            *    98611 98710: gap of      100 bp
            *    98711   111069: contig of 12359 bp in length
            *   111070 111169: gap of      100 bp
            *   111170   122730: contig of 11561 bp in length
            *   122731 122830: gap of      100 bp
            *   122831   135722: contig of 12892 bp in length
            *   135723 135822: gap of      100 bp
            *   135823   150963: contig of 15141 bp in length
            *   150964 151063: gap of      100 bp
            *   151064   164434: contig of 13371 bp in length.
FEATURES             Location/Qualifiers
     source          1. .164434
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-658A5"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1416
                     /note="assembly_fragment"
     misc_feature    1517. .3580
                     /note="assembly_fragment"
     misc_feature    3681. .4383
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    4484. .5844
```

```
                     /note="assembly_fragment"
     misc_feature    5945. .7574
                     /note="assembly_fragment"
     misc_feature    7675. .9988
                     /note="assembly_fragment"
     misc_feature    10089. .13426
                     /note="assembly_fragment"
     misc_feature    13527. .15573
                     /note="assembly_fragment"
     misc_feature    15674. .18974
                     /note="assembly_fragment"
     misc_feature    19075. .21809
                     /note="assembly_fragment"
     misc_feature    21910. .25085
                     /note="assembly_fragment"
     misc_feature    25186. .29425
                     /note="assembly_fragment"
     misc_feature    29526. .34133
                     /note="assembly_fragment"
     misc_feature    34234. .40649
                     /note="assembly_fragment"
     misc_feature    40750. .45891
                     /note="assembly_fragment"
     misc_feature    45992. .53180
                     /note="assembly_fragment"
     misc_feature    53281. .62817
                     /note="assembly_fragment"
     misc_feature    62918. .70247
                     /note="assembly_fragment"
     misc_feature    70348. .78754
                     /note="assembly_fragment"
     misc_feature    78855. .86840
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    86941. .98610
                     /note="assembly_fragment"
     misc_feature    98711. .111069
                     /note="assembly_fragment"
     misc_feature    111170. .122730
                     /note="assembly_fragment"
     misc_feature    122831. .135722
                     /note="assembly_fragment"
     misc_feature    135823. .150963
                     /note="assembly_fragment"
     misc_feature    151064. .164434
                     /note="assembly_fragment"
BASE COUNT    50098 a  34031 c  34547 g  43240 t   2518 others
ORIGIN


alignment_scores:
           Quality: 2257.00            Length:   1735
             Ratio:  3.946              Gaps:      9
 Percent Similarity:  32.968   Percent Identity: 32.795

alignment_block:
 US-09-927-267-1 x AC036216   ..

 Align seg 1/1  to: AC036216  from: 1  to: 164434

       1 MetSerGlnAspThrLysValLysThrThrGluSerSerProProAlaPr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  156640 ATGAGCCAGGACACCAAAGTGAAGACAACAGAGTCCAGTCCCCCAGCCCC 156689

      17 oSerLysAla..................................... 20
         |||||||||
  156690 ATCCAAGGCCAGGTGAGAAGTCCTGGTCCCTTGTGTGGGATCTCTCCTCA 156739

      20 ................................................ 20

  156740 TTCCTCTTGGTGCCCCAGTCACAACTACAGCTTTGAATGCCTGGTGAATA 156789

      21 .....................Arg.LysLeuLeuProValLeuAspP 29
                                 ::: ||||||||||||||||||||
  156790 AATGAAGCAAGACTTTCTTTCTTACAGGAAGTTGCTGCCTGTCCTGGACC 156839

      29 roSerGlyAspTyrTyrTyrTrpTrpLeuAsnThrMetValPheProVal 45
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  156840 CATCTGGGGATTACTACTACTGGTGGCTGAACACAATGGTCTTCCCAGTC 156889

      46 MetTyrAsnLeuIleIleLeuVal..................... 53
         |||||||||||||||||||||||
  156890 ATGTATAACCTCATCATCCTCGTGTGCAGGTATGGCAGCGGTGCTAAGGG 156939

      53 ................................................ 53

  156940 AGGGGCTGGAAGCCAAAAAGAGGACTAAAGAGAGGTCAAGGAGAAGGGCA 156989

      53 ................................................ 53

  156990 GACCCTTGGTGGGGCAGGAGGAGCAATTCCCATGGGAGGGCCTGAGGCAG 157039

      53 ................................................ 53

  157040 AGGGTTAAGGGCCCTGGGGAGACGCCTCGCACACAGAGGGTGCCCTTAAT 157089

      54 .....................CysArgAlaCysPheProAspLeuGlnH 63
                               ||||||||||||||||||||||||||||
  157090 TCAATCATGCTTAACCCTGCCCTGCAGAGCCTGCTTCCCCGACTTGCAGC 157139

      63 isGlyTyrLeuValAlaTrpLeuValLeuAspTyrThrSerAspLeuLeu 79
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  157140 ACGGTTATCTGGTGGCCTGGTTGGTGCTGGACTACACGAGTGACCTGCTA 157189

      80 TyrLeuLeuAspMetValValArgPhe..His..................  89
         ||||||||||||||||||||||||||  |||
  157190 TACCTACTAGACATGGTGGTGCGCTTCCACACAGGTCAGTGGGCTTCTAG 157239

      89 ................................................ 89

  157240 GAATGACCCTTTGTCCCACATTCCCTTCCTAAAGATAGCCACTTAAGAAG 157289

      89 ................................................ 89

  157290 TAACAAGAAAGGCACCCCCACCGTGGTAGCACCTTCGCGTGCCTCTATGC 157339

      89 ................................................ 89

  157340 CTGACAGCATCCCAGTGCTCACCCCGGAAAGCCGGGAGCAGAGTTATGCC 157389

      89 ................................................ 89

  157390 TGGCTCCACTCTGTCCTTCAGACAGTCTCCCTGGCCTGCCCTGGGCAGCT 157439

      89 ................................................ 89

  157440 CATGCTCAGCCCAAGCTTGACTACAGCAGGTCCGCTTCCTACCGGCTCCC 157489

      90 .....ThrGlyPheLeuGluGlnGlyIleLeuValValAspLysGlyArg 104
               ||||||||||||||||||||||||||||||||||||||||||||||||
  157490 TCTCCCCAGGATTCTTGGAACAGGGCATCCTGGTGGTGGACAAGGGTAGG 157539

     105 IleSerSerArgTyrValArgThrTrpSerPhePheLeuAspLeuAlaSe 121
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  157540 ATCTCGAGTCGCTACGTTCGCACCTGGAGTTTCTTCTTGGACCTGGCTTC 157589

     121 rLeuMetProThrAspValValTyrValArgLeuGlyProHisThrProT 138
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  157590 CCTGATGCCCACAGATGTGGTCTACGTGCGGCTGGGCCCGCACACACCCA 157639

     138 hrLeuArgLeuAsnArgPheLeuArgAlaProArgLeuPheGluAlaPhe 154
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  157640 CCCTGAGGCTGAACCGCTTTCTCCGCGCGCCCCGCCTCTTCGAGGCCTTC 157689

     155 AspArgThrGluThrArgThrAlaTyrProAsnAlaPheArgIleAlaLy 171
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
157690 GACCGCACAGAGACCCGCACAGCTTACCCAAATGCCTTTCGCATTGCCAA 157739

   171 sLeuMetLeuTyrIlePheValValIleHisTrpAsnSerCysLeuTyrP 188
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
157740 GCTGATGCTTTACATTTTTGTCGTCATCCATTGGAACAGCTGCCTATACT 157789

   188 heAlaLeuSerArgTyrLeuGlyPheGlyArgAspAlaTrpValTyrPro 204
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
157790 TTGCCCTATCCCGGTACCTGGGCTTCGGGCGTGACGCATGGGTGTACCCG 157839

   205 AspProAlaGlnProGlyPheGluArgLeuArgArgGlnTyrLeuTyrSe 221
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
157840 GACCCCGCGCAGCCTGGCTTTGAGCGCCTGCGGCGCCAGTACCTCTATAG 157889

   221 rPheTyrPheSerThrLeuIleLeuThrThrValGlyAspThrProProP 238
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
157890 CTTTTACTTCTCCACGCTGATACTGACTACAGTGGGCGATACACCGCCGC 157939

   238 roAlaArgGluGluGluTyrLeuPheMetValGlyAspPheLeuLeuAla 254
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
157940 CAGCCAGGGAAGAAGAGTACCTCTTCATGGTGGGCGACTTCCTGCTGGCC 157989

   255 ValMetGlyPheAlaThrIleMetGlySerMetSerSerValIleTyrAs 271
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
157990 GTCATGGGTTTCGCCACCATCATGGGTAGCATGAGCTCTGTCATCTACAA 158039

   271 nMetAsnThrAlaAspAlaAlaPheTyrProAspHisAlaLeuValLysL 288
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
158040 CATGAACACTGCAGATGCGGCTTTCTACCCAGATCATGCACTGGTGAAGA 158089

   288 ysTyrMetLysLeuGlnHisValAsnArgLysLeuGluArgArgValIle 304
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
158090 AGTACATGAAGCTGCAGCACGTCAACCGCAAGCTGGAGCGGCGAGTTATT 158139

   305 AspTrp..Tyr..................................... 307
                 ||||||
158140 GACTGGTGAGAAGGCGGGGTTCCAGACCAGGACAGGGACCAGTGTAGGTG 158189

   307 ..................................................... 307

158190 ATGGAACCTGAGGGAGGTAACTGGGTCCTTAGTGCCTGGTGAGCCAGGCA 158239

   307 ..................................................... 307

158240 AGGCTGTCAAAATGTAGCATTCAGCCGTGGGTTTGCTGGCTGGGGCTTGG 158289

   307 ..................................................... 307

158290 AAAAAGGGAACTTCTTTCAACTGAGGGAATCAGGACTTGGGGGAGGGGTA 158339

   307 ..................................................... 307

158340 GGTAAGAGACTGATAGGGAGAGGAGCTCATACTCAAAAAAGGATAATATG 158389

   307 ..................................................... 307

158390 GAGACCAGGGAATGGGAAGTGCCACTGCCTGGTAGGGCTCAGAAGGCTCT 158439

   307 ..................................................... 307

158440 GGAAGGAGTGGGTGAAGAAGGGCAATCCCCCCTGAGAATGGTCAGCAACA 158489

   307 ..................................................... 307

158490 AGATCATTGGCTACCTATGGTGATTCATACTGTGGACTGAACAAAAGGAA 158539

   307 ..................................................... 307

158540 GTGTCACAAAGTTGGGGAAGTAGAGCAGAGGGTCCCCAGAGAGGAGGGGC 158589

   307 ..................................................... 307

158590 CCTTGGTGGCTGAGGTAGCTAAGGGTAGGGGTGGAACAAGACATCACTGA 158639

   307 ..................................................... 307

158640 TTGGTGGGTAGGGTAGAGCCTGATAGGAGGTGAAGGGTTATGTAAAGTGG 158689

   307 ..................................................... 307

158690 TAGAGGTGTCTATGCCCTGGAAAACAGGTAATCCAACTGTCAGGTACTCC 158739

   307 ..................................................... 307

158740 CATGACCCCTGTTAGATCTCAGTATGGTGGGATTCTTGGCTGGAGCTGAG 158789

   307 ..................................................... 307

158790 CTGAGCCCTGAAGGAAGGGATGGAGGCAGGCACATAAAGGGCTGGTAAGG 158839

   308 ..........................GlnHisLeuGlnIleAsnLysL 315
                                ||||||||||||||||||||||
158840 AAATGGCACTGTCCTTACTCTCAGGTATCAGCACCTGCAGATCAACAAGA 158889

   315 ysMetThrAsnGluValAlaIleLeuGlnHisLeuProGluArgLeuArg 331
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
158890 AGATGACCAACGAGGTAGCCATCTTACAGCACTTGCCTGAGCGGCTGCGG 158939

   332 AlaGluValAlaValSerValHisLeuSerThrLeuSerArgValGlnIl 348
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
158940 GCAGAAGTGGCTGTGTCTGTGCACCTGTCCACTCTGAGCCGGGTGCAGAT 158989

   348 ePheGlnAsnCysGluAlaSerLeuLeuGluGluLeuValLeuLysLeuG 365
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
158990 CTTTCAGAACTGTGAGGCCAGCCTGCTGGAGGAGCTGGTGCTGAAGCTGC 159039

   365 lnProGlnThrTyrSerProGlyGluTyrValCysArgLysGlyAspIle 381
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
159040 AGCCCCAGACCTACTCACCAGGTGAATATGTATGCCGCAAAGGAGACATT 159089

   382 GlyGlnGluMetTyrIleIleArgGluGlyGlnLeuAlaValValAlaAs 398
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
159090 GGCCAAGAGATGTACATCATCCGAGAGGGTCAACTGGCCGTGGTGGCAGA 159139

   398 pAspGlyIleThrGlnTyrAlaValLeuGlyAlaGlyLeuTyrPheGlyG 415
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
159140 TGATGGTATCACACAGTATGCTGTGCTCGGTGCAGGGCTCTACTTTGGGG 159189

   415 luIleSerIleIleAsnIle..Lys..................... 422
       ||||||||||||||||||  :::
159190 AGATCAGCATCATCAACATCAAAGGTGGGTATCCCAGTATTTGTTCCAGG 159239

   422 ..................................................... 422

159240 GACAAGGATGGGTGGGGTAGGGGGGAACAGCAGAGCCCAGTGCTGGGACC 159289

   422 ..................................................... 422

159290 AGATAGTACTTCAGGCCTAAACTTCTGATTGAGGAAACCTGGCCCTTCTC 159339

   422 ..................................................... 422

159340 TGAGTCACTAGATGGCTCAGGAGAAAAACAACACATAGAGGATCCATTCC 159389

   422 ..................................................... 422

159390 CTGAGAAGAAGCTGAGCCAAGGCCAGTGAACAGGGTGGGTTCGTGGAAGA 159439

   422 ..................................................... 422

159440 GAGAGTAGACTTGCTCTGTGGATATTCACCTGAAGCACCAATAGACTAGT 159489

   422 ..................................................... 422

159490 GGACAGTGATAGAGATAATAATAATAATAATAGCTAATATTTATTACCTG 159539
```

```
   422 ..................................................  422
159540 CTTACTGCATCCCCTTTTCTAAGTGCTTCACATGATATTAACTCATTACA 159589
   422 ..................................................  422
159590 TTCCCGCAGCAGTTCTATAAGATAAGTACTATTCTTATCCCAATTTTAAA 159639
   422 ..................................................  422
159640 GATAAGGAAATCAAACTCAAAGAAGTTGGGTAATTTGCCCAACGTCGCAC 159689
   422 ..................................................  422
159690 AGGTATACACATGTGGGATGAGTGGAAAAAACAACCTTTTTACAGCCAAG 159739
   422 ..................................................  422
159740 TGGAAAGAGAAAGGTAGAAGAGAAGAAGTTCCCATTTACTGACCATCACC 159789
   422 ..................................................  422
159790 TCTAGGCCAAGCCAGGTACTACCTACTGGCTTGCTTTTTCCTCTCATCAG 159839
   422 ..................................................  422
159840 CCCTGTGCAGTAGGAATTGTTATGCCTAATCACAGGTATTAGGCATGAGG 159889
   422 ..................................................  422
159890 TCGTGCAAGCAGGAAATCCTGCAAATGGGGTAAGCACACTGTCATGGACC 159939
   422 ..................................................  422
159940 GAGAAGCTGGATGTGCTCTGGCTCTCTCAGCATGAGGATTCTACCCTCAG 159989
   422 ..................................................  422
159990 GGATTCTGAAAGCCTCACATCTGTCCTCCAATGATCCCCTCAAGCTCCAC 160039
   422 ..................................................  422
160040 ATCCTGTATGTGGTAGGATGCTGTCTGTTTGACTTTGGACAAGTCATTTC 160089
   422 ..................................................  422
160090 CCTATTTAGGCCTCAGTTTAATCACCTATCCAATGAGGACAATAATAAAA 160139
   422 ..................................................  422
160140 ACCCAATGAGTTATTTGTGAGAATAAATGAGCTATGTGAGTGAGGCATCC 160189
   422 ..................................................  422
160190 AGCGCAGTGCTTGGCATACAGCAAGTGTTCAACAAATAGTAGCTCCCTTT 160239
   422 ..................................................  422
160240 CCCTTGTCCAAGGTCCTTCTAGCTCTAAGACTTTGCAGCATTCAGCCTCA 160289
   422 ..................................................  422
160290 CCTATCTCCCCACCCATTCTCCACCAACATCTTTCTAAACTGCAAAGGAT 160339
   422 ..................................................  422
160340 ATCATACAGCTCCCTGCTTAAAATACTTCAATAGTTTCCTACTGCCCTCA 160389
   422 ..................................................  422
160390 GCATAAGGCCTGAACTTCATGTCATGGCTTATGAGACCCAGTGTGACCTA 160439
```

```
   422 ..................................................  422
160440 GATCTCCCCTCTCTCCACCCTCCCCACACTCTGCGCTTCTCGCACTCTGA 160489
   422 ..................................................  422
160490 ACTGCTTACTATTTTCTGCATCCAACTGACTCTTTTCTACCCCCTTCCCT 160539
   422 ..................................................  422
160540 CTGTTCCTGCTCCTTCCTCTGGCTAGATTGCCCTATCCCCACTTGTCCCC 160589
   422 ..................................................  422
160590 CTCCTCCCCCTAGTTACCTCCTACTCAGTTCAGGTGTATGAGACTGTTCT 160639
   422 ..................................................  422
160640 TGCATTTCTATAAAGAAATATCTGAGACTGGATAATTTATAAAGAAAAGA 160689
   422 ..................................................  422
160690 GGTTTAATTGGCTCACAGTTCTTCAGGATTTACGGGAAGCATGGTGCTGG 160739
   422 ..................................................  422
160740 CATTTGCTCAGCTTCTAGGGAGGCTTCAAGAAGCTTATAATTGTGGCAGA 160789
   422 ..................................................  422
160790 AGGCAAAGGGGGAGCAGGCATGTCACACGGTGAAAGCAGGAGCAAGGGTT 160839
   422 ..................................................  422
160840 GGGGGAGGTGCCACACTTTCAAACAACCAAAACAACCAGCTCTCGACTCA 160889
   422 ..................................................  422
160890 CTAACTCAAAGACAGCATCAAGCCATGAAGGATTCGCCCCCATAATCCAA 160939
   422 ..................................................  422
160940 TCACCTCCCACCAGGTCCCACCCCCAACATTGGGGATTACATTTCACATG 160989
   422 ..................................................  422
160990 AGATTTGGGCAGAGACAAATATTCAAATTATATCTTAGGACATCCCTTCC 161039
   422 ..................................................  422
161040 TCCAGACTTCCCTAAATTCCCTGCCTTACAGTTTGGTAGGGGCTCTTTGC 161089
   422 ..................................................  422
161090 CTACCTTTCCACAGCACCTGATGAACATGTCTTCACTGCACCAGCCATAC 161139
   422 ..................................................  422
161140 TGTTATATAACTATTCCAATATATGTGTATCTCCTCTAGACTGTGAATTA 161189
   422 ..................................................  422
161190 TTTGAGGCAGGTCACTGATACCTACCCAGCATGGAGCCTGGTCCCCAGTA 161239
   422 ..................................................  422
161240 TGTTACTAAATGAAAGAATTAATGAGGCAGAAGGAGAGGCTCAGAAGCAC 161289
   422 ..................................................  422
161290 AAATATGGAGGTGAAGGTCCTGGTTCAGGAGTGAAATGCCACCTCCTCAC 161339
   423 .....................................GlyAsnMet  425
```

```
                                                          |||||||||
  161340 CCTCCTACTAACTGTCCTCCCATCTCTGCCCATGAGCCACAGGGAACATG 161389

     426 SerGlyAsnArgArgThrAlaAsnIleLysSerLeuGlyTyrSerAspLe 442
         |||||||||||||||||||||||||||||||||||||||||||||||||||
  161390 TCTGGGAACCGCCGCACAGCCAACATCAAGAGCCTAGGTTATTCAGACCT 161439

     442 uPheCysLeuSerLysGluAspLeuArgGluValLeuSerGluTyrProG 459
         |||||||||||||||||||||||||||||||||||||||||||||||||||
  161440 ATTCTGCCTGAGCAAGGAGGACCTGCGGGAGGTGCTGAGCGAGTATCCAC 161489

     459 lnAlaGlnThrIleMetGluGluLysGlyArgGluIleLeuLeuLysMet 475
         |||||||||||||||||||||||||||||||||||||||||||||||||||
  161490 AAGCACAGACCATCATGGAGGAGAAAGGACGTGAGATCCTGCTGAAAATG 161539

     476 AsnLysLeuAspValAsnAlaGluAlaAlaGluIleAlaLeuGlnGluAl 492
         |||||||||||||||||||||||||||||||||||||||||||||||||||
  161540 AACAAGTTGGACGTGAATGCTGAGGCAGCTGAGATCGCCCTGCAGGAGGC 161589

     492 aThrGluSerArgLeuArgGlyLeuAspGlnGlnLeuAspAspLeuGlnT 509
         |||||||||||||||||||||:::|||||||||||||||||||||||||||
  161590 CACAGAGTCCCGGCTACGAGGCATAGACCAGCAGCTGGATGATCTACAGA 161639

     509 hrLysPheAlaArgLeuLeuAlaGluLeuGluSerSerAlaLeuLysIle 525
         |||||||||||||||||||||||||||||||||||||||||||||||||||
  161640 CCAAGTTTGCTCGCCTCCTGGCTGAGCTGGAGTCCAGCGCACTTAAGATT 161689

     526 AlaTyrArgIleGluArgLeuGluTrpGlnThrArgGluTrpProMetPr 542
         |||||||||||||||||||||||||||||||||||||||||||||||||||
  161690 GCTTACCGCATTGAACGGCTGGAGTGGCAGACTCGAGAGTGGCCAATGCC 161739

     542 oGluAspLeuAlaGluAlaAspAspGluGlyGluProGluGluGlyThrS 559
         |||||||||||||||||||||||||||||||   ||||||||||||||||
  161740 CGAGGACCTGGCTGAGGCTGATGACGAGGGTGTGCCTGAGGAGGGAACTT 161789

     559 erLysAspGluGluGlyArgAlaSerGlnGluGlyProProGlyProGlu 575
         |||||||||||||||||||||||||||||||||||||||||||||||||||
  161790 CCAAAGATGAAGAGGGCAGGGCCAGCCAGGAGGGACCCCCAGGTCCAGAG 161839

seq_name: gb_ro:MMU243933

seq_documentation_block:
LOCUS       MMU243933               2781 bp    mRNA    linear   ROD 25-MAY-2000
DEFINITION  Mus musculus mRNA for cone photoreceptor cyclic nucleotide-gated
            channel alpha-subunit.
ACCESSION   AJ243933
VERSION     AJ243933.1  GI:7688040
KEYWORDS    CNG3 gene; cyclic nucleotide-gated channel alpha-subunit.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2781)
  AUTHORS   Hirano,A.A., Hack,I., Wassle,H. and Duvoisin,R.M.
  TITLE     Cloning and immunocytochemical localization of a cyclic
            nucleotide-gated channel alpha-subunit to all cone photoreceptors
            in the mouse retina
  JOURNAL   J. Comp. Neurol. 421 (1), 80-94 (2000)
  MEDLINE   20273944
REFERENCE   2  (bases 1 to 2781)
  AUTHORS   Hirano,A.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-1999) Hirano A.A., Neuroanatomie,
            Max-Planck-Institut fuer Hirnforschung, Deutschordenstrasse 46,
            D-60528 Frankfurt a.M., GERMANY
FEATURES             Location/Qualifiers
     source          1. .2781
                     /organism="Mus musculus"
                     /plasmid="pNC-2"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /tissue_type="retina"
                     /dev_stage="adult"
     gene            380. .2275
                     /gene="CNG3"
     CDS             380. .2275
                     /gene="CNG3"
                     /function="phototransduction in cone photoreceptors"
                     /codon_start=1
                     /evidence=experimental
                     /product="cyclic nucleotide-gated channel alpha-subunit"
                     /protein_id="CAB89685.1"
                     /db_xref="GI:7688041"
                     /translation="MAKVNTQCSQPSPTQLSIKNADRDLDHVENGLGRVSRLIISIRA
                     WASRHLHDEDQTPDSFLDRFHGSELKEVSTRESNAQPNPGEQKPPDGGEGRKEEPIVV
                     DPSSNIYCRWLTAIALPVFYNWCLLVCRACFDELQSEHLTLWLVLDYSADVLYVLDML
                     VRARTGFLEQGLMVRDTKRLWKHYTKTLHFKLDILSLIPTDLAYLKLGVNYPELRFNR
                     LLKFSRLFEFFDRTETRTNYPNVFRIGNLVLYTLIIIHWNACIYFAISKFIGFGTDSW
                     VYPNTSKPEYARLSRKYIYSLYWSTLTLTTIGETPPPVKDEEYLFVVIDFLVGILIFA
                     TIVGNVGSMISNMNAPRVEFQAKIDSVKQYMQFRKVTKDLETRVIRWFDYLWANRKTV
                     DEKEVLKNLPDKLKAEIAINVHLDTLKKVRIFQDCEAGLLVELVLKLRPTVFSPGDYI
                     CKKGDIGREMYIIKEGKLAVVADDGVTQFVVLSDGSYFGEISILNIKGSKSGNRRTAN
                     IRSIGYSDLFCLSKDDLMEALTEYPDAKRALEEKGRQILMKDNLIDEDLVAARVDTRD
                     VEEKVEYLESSLDILQTRFARLLAEYSASQMKLKQRLTRLESQMNRRCCGFSPDRENS
                     EDASKTD"
BASE COUNT      672 a    707 c    743 g    659 t
ORIGIN


alignment_scores:
            Quality: 1580.50             Length:    587
              Ratio:  3.363               Gaps:      6
 Percent Similarity:  80.068   Percent Identity: 52.470

alignment_block:
 US-09-927-267-1 x MMU243933   ..

 Align seg 1/1  to: MMU243933  from: 1  to: 2781

       2 SerGlnAspThrLysValLysThrThrGluSerSer..............  13
         |||:::      :::|||:::|||   |||||||:::
     578 TCTGAGCTTAAGGAAGTCTCCACCCGGGAAAGCAATGCCCAGCCCAACCC 627

      14 .............ProProAlaProSerLysAlaArgLysLeuLeuPro.  25
                      ||||||      ::::::::::||||||      |||
     628 AGGAGAACAGAAGCCACCAGACGGAGGGGAAGGCAGGAAGGAGGAGCCCA 677

      26 ..ValLeuAspProSerGlyAspTyrTyrTyrTrpTrpLeuAsnThrMet  41
           |||:::||||||||||:::::   |||      ||||||:::::::::
     678 TTGTGGTGGACCCCTCCAGCAACATCTACTGCCGCTGGCTGACTGCCATC 727

      42 ValPheProValMetTyrAsnLeuIleIleLeuValCysArgAlaCysPh  58
         ::::::||||||:::||||||      :::||||||||||||||||||||
     728 GCGCTCCCAGTCTTCTATAACTGGTGTCTACTTGTATGCAGGGCCTGTTT 777

      58 eProAspLeuGlnHisGlyTyrLeuValAlaTrpLeuValLeuAspTyrT  75
         |   :::||||||      :::|||:::   ||||||||||||||||||:
     778 TGATGAGCTACAATCAGAGCACCTGACGCTGTGGCTGGTCTTGGACTACT 827

      75 hrSerAspLeuLeuTyrLeuLeuAspMetValValArgPheHisThrGly  91
         :::::|||:::||||||:::|||||||||:::||||||   :::||||||
     828 CTGCAGATGTCCTTTATGTTTTAGACATGCTGGTTCGAGCCCGGACAGGT 877

      92 PheLeuGluGlnGlyIleLeuValValAspLysGlyArgIleSerSerAr 108
         ||||||||||||||:::::|||   |||      |||:::   :::::
     878 TTCCTTGAGCAAGGCCTAATGGTCAGAGATACCAAGAGGCTGTGGAAACA 927

     108 gTyrValArgThrTrpSerPhePheLeuAspLeuAlaSerLeuMetProT 125
         :|||::::::|||      |||   ||||||:::   ||||||:::||||
     928 TTACACAAAGACCTTGCACTTCAAGCTGGACATCCTGTCTCTCATCCCCA 977

     125 hrAspValValTyrValArgLeuGlyProHisThrProThrLeuArgLeu 141
         |||||:::::::|||::::::||||||   :::   |||   |||||||::
     978 CAGACCTGGCTTATTTAAAGTTGGGCGTGAACTACCCAGAACTGAGGTTC 1027

     142 AsnArgPheLeuArgAlaProArgLeuPheGluAlaPheAspArgThrGl 158
```

```
1028 AACCGCCTCCTGAAGTTCTCTCGGCTCTTTGAATTCTTTGACCGCACAGA 1077

 158 uThrArgThrAlaTyrProAsnAlaPheArgIleAlaLysLeuMetLeuT 175
1078 GACAAGGACCAACTACCCCAACGTGTTCAGGATAGGGAACTTGGTCCTTT 1127

 175 yrIlePheValValIleHisTrpAsnSerCysLeuTyrPheAlaLeuSer 191
1128 ACACACTCATCATCATCCACTGGAATGCCTGCATCTACTTTGCCATTTCC 1177

 192 ArgTyrLeuGlyPheGlyArgAspAlaTrpValTyrProAspProAlaGl 208
1178 AAGTTCATTGGTTTTGGGACAGACTCCTGGGTCTATCCAAACACCTCCAA 1227

 208 nProGlyPheGluArgLeuArgArgGlnTyrLeuTyrSerPheTyrPheS 225
1228 GCCGGAGTATGCACGACTCTCCCGGAAGTACATTTACAGTCTCTACTGGT 1277

 225 erThrLeuIleLeuThrThrValGlyAspThrProProProAlaArgGlu 241
1278 CCACCCTGACCCTGACCACCATTGGTGAGACCCCACCCCCCGTGAAAGAT 1327

 242 GluGluTyrLeuPheMetValGlyAspPheLeuLeuAlaValMetGlyPh 258
1328 GAAGAGTATCTCTTTGTGGTCATCGACTTCCTGGTGGGTATCCTGATCTT 1377

 258 eAlaThrIleMetGlySerMetSerSerValIleTyrAsnMetAsnThrA 275
1378 TGCCACCATAGTGGGCAACGTAGGCTCCATGATTTCCAACATGAACGCTC 1427

 275 laAspAlaAlaPheTyrProAspHisAlaLeuValLysLysTyrMetLys 291
1428 CCCGGGTAGAGTTCCAGGCTAAGATCGACTCCGTCAAGCAGTACATGCAG 1477

 292 LeuGlnHisValAsnArgLysLeuGluArgArgValIleAspTrpTyrGl 308
1478 TTCCGGAAGGTAACCAAGGACTTGGAGACTCGGGTTATCCGGTGGTTTGA 1527

 308 nHisLeuGlnIleAsnLysLysMetThrAsnGluValAlaIleLeuGlnH 325
1528 CTATCTGTGGGCCAACAGGAAGACGGTGGATGAAAAGGAAGTGCTCAAAA 1577

 325 isLeuProGluArgLeuArgAlaGluValAlaValSerValHisLeuSer 341
1578 ACCTCCCGGACAAGCTGAAGGCTGAGATCGCCATCAACGTGCACCTGGAC 1627

 342 ThrLeuSerArgValGlnIlePheGlnAsnCysGluAlaSerLeuLeuGl 358
1628 ACGCTGAAGAAGGTCCGAATCTTCCAGGACTGCGAGGCGGGGCTGCTGGT 1677

 358 uGluLeuValLeuLysLeuGlnProGlnThrTyrSerProGlyGluTyrV 375
1678 GGAGCTGGTGCTGAAGCTCCGCCCCACTGTGTTCAGCCCCGGGGACTACA 1727

 375 alCysArgLysGlyAspIleGlyGlnGluMetTyrIleIleArgGluGly 391
1728 TATGCAAAAAGGGGGACATTGGAAGAGAGATGTACATCATCAAGGAGGGC 1777

 392 GlnLeuAlaValValAlaAspAspGlyIleThrGlnTyrAlaValLeuGl 408
1778 AAGCTGGCTGTGGTGGCTGATGACGGGGTCACCCAGTTTGTGGTCCTCAG 1827

 408 yAlaGlyLeuTyrPheGlyGluIleSerIleIleAsnIleLysGlyAsnM 425
1828 TGACGGCAGTTACTTTGGGGAGATTAGCATCTTAAACATTAAGGGGAGCA 1877

 425 etSerGlyAsnArgArgThrAlaAsnIleLysSerLeuGlyTyrSerAsp 441
1878 AGTCTGGGAACCGCAGGACGGCCAACATCAGGAGTATCGGCTACTCAGAC 1927

 442 LeuPheCysLeuSerLysGluAspLeuArgGluValLeuSerGluTyrPr 458
```

```
1928 CTGTTCTGCCTCTCCAAGGATGACCTAATGGAAGCCCTCACCGAGTACCC 1977

 458 oGlnAlaGlnThrIleMetGluGluLysGlyArgGluIleLeuLeuLysM 475
1978 AGACGCTAAGAGGGCTCTGGAGGAAAAGGGCCGTCAGATTCTGATGAAGG 2027

 475 etAsnLysLeuAspValAsnAlaGluAlaAlaGluIleAlaLeuGlnGlu 491
2028 ACAACCTAATTGATGAGGACCTAGTCGCGGCCAGGGTAGATACCAGGGAC 2077

 492 AlaThrGluSerArgLeuArgGlyLeuAspGlnGlnLeuAspAspLeuGl 508
2078 ...GTTGAGGAGAAGGTGGAGTACCTGGAGTCGTCCCTGGACATCCTGCA 2124

 508 nThrLysPheAlaArgLeuLeuAlaGluLeuGluSerSerAlaLeuLysI 525
2125 GACGAGGTTTGCCCGACTCCTGGCTGAGTACAGTGCCTCCCAGATGAAGC 2174

 525 leAlaTyrArgIleGluArgLeuGluTrpGlnThrArgGluTrpProMet 541
2175 TGAAACAGCGGCTCACTCGGCTGGAGAGCCAG.............ATG 2209

 542 ProGluAspLeuAlaGluAlaAspAspGluGlyGluProGluGluGlyTh 558
2210 AACAGGAGGTGTTGTGGCTTCTCACCTGACAGGGAGAATTCCGAGGATGC 2259

 558 rSerLysAspGlu.Glu.........GlyArgAlaSerGlnGluGlyPro 571
2260 TTCAAAGACTGACTGAAAATGCAGGTGGCTCTGCCTCCTGCTCCCCCCCC 2309

 572 ProGlyPro 574
2310 CCCCCCCCG 2318
```

```
seq_name: gb_om:BTCNGCHAN

seq_documentation_block:
LOCUS       BTCNGCHAN              2641 bp    mRNA    linear   MAM 20-JUL-1995
DEFINITION  B.taurus mRNA for alpha subunit of cone and testis CNG channel.
ACCESSION   X89600
VERSION     X89600.1  GI:908823
KEYWORDS    alpha-subunit; CNG channel.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 2641)
  AUTHORS   Weyand,I., Godde,M., Frings,S., Weiner,J., Muller,F.,
            Altenhofen,W., Hatt,H. and Kaupp,U.B.
  TITLE     Cloning and functional expression of a cyclic-nucleotide-gated
            channel from mammalian sperm
  JOURNAL   Nature 368 (6474), 859-863 (1994)
  MEDLINE   94211295
REFERENCE   2  (bases 1 to 2641)
  AUTHORS   Weyand,I.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUL-1995) I. Weyand, Forschungszentrum Julich,
            Institut fur Biologische Informationsverarbeitung, Postfach 1913,
            52425 Julich, FRG
FEATURES             Location/Qualifiers
     source          1. .2641
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
     CDS             151. .2271
                     /codon_start=1
                     /product="alpha subunit of CNG-channel expressed in bovine
                     testis and retinal cone"
                     /protein_id="CAA61759.1"
                     /db_xref="GI:908824"
                     /db_xref="SWISS-PROT:Q29441"
                     /translation="MAKISTQYSHPTRTHPSVRTMDRDLDCIENGLSRTHLPCEETSS
                     ELQEGIAMETRGLAESRQSSFTSQGPTRLSRLIISLRAWSARHLHQEDQRPDSFLERF
```

```
                     RGAELQEVSSRESHVQFNVGSQEPPDRGRSAWPLARNNTNTCNNSEKDDKAKKEEKEK
                     KEEKKENPKKEEKKKDSVVMDPSSNMYYHWLTVIAVPVFYNWCLLVCRACFDELQSEH
                     LMLWLVLDYSADILYGMDMLVRARTGFLEQGLMVMDASRLWKHYTQTLHFKLDVLSLV
                     PTDLAYFKLGMNYPELRFNRLLKLARLFEFFDRTETRTNYPNMFRIGNLVLYILIIIH
                     WNACIYFAISKFIGFGTDSWVYPNVSNPEYGRLSRKYIYSLYWSTLTLTTIGETPPPV
                     KDEEYLFVVIDFLVGVLIFATIVGNVGSMISNMNASRAEFQAKIDSIKQYMQFRKVTK
                     DLETRVIRWFDYLWANKKTVDEKEVLKSLPDKLKAEIAINVHLDTLRKVRIFQDCEAG
                     LLVELVLKLRPAVFSPGDYICKKGDIGREMYIIKEGKLAVVAEDGITQFVVLGDGSYF
                     GEISILNIKGSKSGNRRTANIRSIGYSDLFCLSKDDLMEALTEYPEAKKALEEKGRQI
                     LMKDNLIDEELAKAGADPKDIEEKVEHLETSLDSLQTRFARLLAEYNATQMKVKQRLS
                     QLESQVKMGLPPDGDAPQTEASQP"
BASE COUNT      623 a    727 c    762 g    529 t
ORIGIN


alignment_scores:
            Quality: 1579.00              Length:    564
              Ratio:  3.486                Gaps:      3
 Percent Similarity:  80.319   Percent Identity: 52.660

alignment_block:
 US-09-927-267-1 x BTCNGCHAN   ..

 Align seg 1/1  to: BTCNGCHAN  from: 1  to: 2641

      3 GlnAspThrLysValLysThrThrGluSerSerPro..............  14
        :::|||   |||:::|||      |||::::::
    595 AAGGATGACAAGGCGAAAAAGGAGGAGAAAGAGAAAAAGGAAGAGAAAAA  644

     15 .......ProAlaProSerLysAlaArgLysLeuLeuProValLeuAspP  29
              |||      ::::::   :::|||          |||:::||||
    645 GGAGAACCCCAAGAAAGAGGAGAAGAAGAAGGACAGCGTTGTGATGGACC  694

     29 roSerGlyAspTyrTyrTyrTrpTrpLeuAsnThrMetValPheProVal  45
        |||||::::::   ||||||   ||||||:::::::::::   ||||||
    695 CTTCCAGCAACATGTACTACCACTGGCTGACTGTCATCGCCGTGCCTGTC  744

     46 MetTyrAsnLeuIleIleLeuValCysArgAlaCysPheProAspLeuGl  62
        :::||||||      :::|||||||||||||||||||||   :::||||||
    745 TTCTACAACTGGTGTCTGCTCGTGTGCAGGGCCTGTTTCGATGAGCTCCA  794

     62 nHisGlyTyrLeuValAlaTrpLeuValLeuAspTyrThrSerAspLeuL  79
        |      :::|||:::   |||||||||||||||||||:::::::|||:::|
    795 GTCCGAGCACCTGATGCTTTGGCTGGTCCTGGACTACTCAGCAGACATCC  844

     79 euTyrLeuLeuAspMetValValArgPheHisThrGlyPheLeuGluGln  95
        |||||   :::|||||||:::||||||   :::|||||||||||||||||
    845 TCTATGGCATGGATATGCTGGTCCGAGCCCGGACAGGCTTCCTGGAGCAA  894

     96 GlyIleLeuValValAspLysGlyArgIleSerSerArgTyrValArgTh 112
        |||:::::|||:::|||   :::|||:::   :::::::|||:::::::||
    895 GGCCTGATGGTCATGGACGCCAGCCGGCTGTGGAAGCACTACACACAGAC  944

    112 rTrpSerPhePheLeuAspLeuAlaSerLeuMetProThrAspValValT 129
        |      |||   ||||||:::   ||||||:::|||||||||:::::::|
    945 CTTGCACTTCAAGCTGGACGTGTTGTCCCTGGTGCCCACAGACCTGGCTT  994

    129 yrValArgLeuGlyProHisThrProThrLeuArgLeuAsnArgPheLeu 145
        ||   :::|||||||   :::   |||   ||||||:::|||||||:::|||
    995 ATTTTAAGCTGGGCATGAACTACCCAGAACTGAGGTTCAACCGCCTCCTG 1044

    146 ArgAlaProArgLeuPheGluAlaPheAspArgThrGluThrArgThrAl 162
        :::      ||||||||||||   ||||||||||||||||||||||
   1045 AAGTTGGCCCGGCTCTTCGAATTCTTTGACCGCACGGAGACAAGGACCAA 1094

    162 aTyrProAsnAlaPheArgIleAlaLysLeuMetLeuTyrIlePheValV 179
         |||||||||   |||||||||:::::::|||:::||||||||:::::::
   1095 CTACCCCAACATGTTCAGGATCGGGAACTTGGTCTTGTACATCCTCATCA 1144

    179 alIleHisTrpAsnSerCysLeuTyrPheAlaLeuSerArgTyrLeuGly 195
        ::|||||||||||||:::|||:::||||||||||:::|||:::::::|||
   1145 TCATCCACTGGAATGCCTGCATCTACTTTGCCATTTCCAAGTTCATTGGT 1194

    196 PheGlyArgAspAlaTrpValTyrProAspProAlaGlnProGlyPheGl 212
        ||||||   |||:::||||||||||||||:::   :::::::|||   :::
   1195 TTCGGGACCGACTCCTGGGTCTACCCAAACGTCTCCAACCCAGAGTATGG 1244

    212 uArgLeuArgArgGlnTyrLeuTyrSerPheTyrPheSerThrLeuIleL 229
         ||||||   |||:::|||:::||||||:::|||:::|||||||||   |
   1245 GCGCCTCTCCAGAAAGTACATTTACAGTCTCTACTGGTCCACCTTGACCC 1294

    229 euThrThrValGlyAspThrProProProAlaArgGluGluGluTyrLeu 245
        ||||||||:::|||:::||||||||||||||:::::::::|||||||||||
   1295 TGACCACCATTGGGGAGACCCCGCCCCCCGTAAAAGACGAGGAATATCTC 1344

    246 PheMetValGlyAspPheLeuLeuAlaValMetGlyPheAlaThrIleMe 262
        |||:::|||   |||||||||:::::|||:::   ||||||||||||::
   1345 TTTGTGGTCATCGACTTCCTGGTGGGCGTCCTGATTTTTGCCACCATCGT 1394

    262 tGlySerMetSerSerValIleTyrAsnMetAsnThrAlaAspAlaAlaP 279
        :|||::::::::::|||:::|||   |||||||||:::::   |||   |
   1395 GGGCAACGTGGGCTCCATGATCTCAAACATGAATGCTTCACGGGCCGAGT 1444

    279 heTyrProAspHisAlaLeuValLysLysTyrMetLysLeuGlnHisVal 295
        ||                 :::|||:::||||||:::::::::   |||
   1445 TCCAGGCCAAGATCGATTCCATCAAGCAGTACATGCAGTTCCGCAAGGTG 1494

    296 AsnArgLysLeuGluArgArgValIleAspTrpTyrGlnHisLeuGlnIl 312
        ::::::   ||||||   |||||||||   |||::::::::::|||
   1495 ACCAAGGACTTGGAGACACGGGTGATCCGCTGGTTCGACTACCTGTGGGC 1544

    312 eAsnLysLysMetThrAsnGluValAlaIleLeuGlnHisLeuProGluA 329
         |||||||||   :::::::|||      :::|||:::   ||||||::::
   1545 CAATAAGAAGACAGTGGATGAGAAGGAGGTGCTCAAGAGCCTCCCCGACA 1594

    329 rgLeuArgAlaGluValAlaValSerValHisLeuSerThrLeuSerArg 345
        ::|||:::||||||:::|||::::::||||||||||:::||||||   :::
   1595 AGCTGAAGGCCGAGATCGCCATCAACGTGCACCTGGACACCCTGAGGAAG 1644

    346 ValGlnIlePheGlnAsnCysGluAlaSerLeuLeuGluGluLeuValLe 362
        |||:::||||||||:::||||||||:::||||||   ||||||||||
   1645 GTCCGAATCTTCCAGGACTGCGAGGCGGGGCTGCTGGTGGAGCTGGTGCT 1694

    362 uLysLeuGlnProGlnThrTyrSerProGlyGluTyrValCysArgLysG 379
        |||||||:::|||   ::::::||||||||:::|||:::|||:::||||
   1695 GAAGCTGCGGCCGGCAGTGTTCAGCCCCGGGGACTACATCTGCAAGAAGG 1744

    379 lyAspIleGlyGlnGluMetTyrIleIleArgGluGlyGlnLeuAlaVal 395
        ||||||||||:::||||||||||||||:::||||||:::||||||||||
   1745 GGGACATCGGGAGGGAGATGTACATTATCAAGGAGGGCAAGCTGGCCGTG 1794

    396 ValAlaAspAspGlyIleThrGlnTyrAlaValLeuGlyAlaGlyLeuTy 412
        ||||||:::||||||||||||||||:::::|||||||||   |||   ||
   1795 GTGGCCGAGGACGGCATCACCCAGTTCGTGGTCCTCGGCGACGGGAGTTA 1844

    412 rPheGlyGluIleSerIleIleAsnIleLysGlyAsnMetSerGlyAsnA 429
        ||||||||||||||||||:::||||||||||||:::   ||||||||||
   1845 TTTCGGGGAGATCAGCATCTTGAACATCAAGGGGAGCAAGTCCGGGAACC 1894

    429 rgArgThrAlaAsnIleLysSerLeuGlyTyrSerAspLeuPheCysLeu 445
        ||||||||||||||||||:::|||:::||||||||||||||||||||||||
   1895 GCCGCACGGCCAACATCAGGAGCATCGGCTACTCGGACCTGTTCTGCCTC 1944

    446 SerLysGluAspLeuArgGluValLeuSerGluTyrProGlnAlaGlnTh 462
        ||||||:::||||||   |||:::|||:::||||||||||:::|||:::
   1945 TCCAAGGACGACCTGATGGAGGCGCTCACCGAGTACCCCGAGGCCAAGAA 1994

    462 rIleMetGluGluLysGlyArgGluIleLeuLeuLysMetAsnLysLeuA 479
            :::|||||||||||||||||:::|||||||:::|||   |||   :::|
   1995 GGCGCTGGAGGAGAAAGGGCGGCAGATCCTCATGAAGGACAACCTGATCG 2044

    479 sp..............ValAsnAlaGluAlaAlaGluIleAlaLeuGln 490
        ||                ::::::|||:::       :::|||
   2045 ACGAGGAGCTGGCCAAGGCCGGGGCAGACCCCAAGGACATC......... 2085

    491 GluAlaThrGluSerArgLeuArgGlyLeuAspGlnGlnLeuAspAspLe 507
```

```
                 |||:::::::::::::   |||:::   :::||||||:::||
    2086 .........GAGGAGAAGGTGGAGCACCTCGAGACCTCCCTGGACTCCCT 2126

     507 uGlnThrLysPheAlaArgLeuLeuAlaGluLeuGluSerSerAlaLeuL 524
         |||||||:::||||||||||||||||||||||   :::::::::   :::|
    2127 GCAGACCAGGTTTGCGCGGCTCCTGGCTGAGTACAACGCCACCCAGATGA 2176

     524 ysIleAlaTyrArgIleGluArgLeuGluTrpGlnThrArgGluTrpPro 540
         ||:::      |||::::::::::||||||   |||::::::
    2177 AGGTGAAGCAGCGGCTCAGCCAGTTGGAAAGCCAGGTGAAGATGGGCCTC 2226

     541 MetProGluAspLeuAlaGluAlaAspAspGluGlyGluPro 554
           |||:::      |||         :::   ::::::|||
    2227 CCGCCTGATGGCGATGCTCCGCAAACTGAGGCCAGTCAGCCC 2268

seq_name: gb_om:BTCNG3

seq_documentation_block:
LOCUS       BTCNG3                  3998 bp    mRNA    linear   MAM 19-MAY-1994
DEFINITION  B.taurus CNG3 mRNA for cyclic nucleotide-gated channel.
ACCESSION   X76485
VERSION     X76485.1  GI:488728
KEYWORDS    cyclic nucleotide-gated channel.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 3998)
  AUTHORS   Biel,M., Zong,X., Distler,M., Bosse,E., Klugbauer,N., Murakami,M.,
            Flockerzi,V. and Hofmann,F.
  TITLE     Another member of the cyclic nucleotide-gated channel family,
            expressed in testis, kidney, and heart
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 91 (9), 3505-3509 (1994)
  MEDLINE   94224768
REFERENCE   2  (bases 1 to 3998)
  AUTHORS   Hofmann,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-NOV-1993) F. Hofmann, Inst. f. Pharmakologie Und
            Toxikologie, Technische Univ. Muenchen, Biedersteiner Str. 29,
            80802 Muenchen, FRG
FEATURES             Location/Qualifiers
     source          1. .3998
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /tissue_type="kidney"
                     /clone_lib="plasmid vector pSport (BRL)"
                     /dev_stage="adult"
     gene            273. .2393
                     /gene="CNG3"
     CDS             273. .2393
                     /gene="CNG3"
                     /codon_start=1
                     /product="cyclic nucleotide-gated channel 3"
                     /protein_id="CAA54023.1"
                     /db_xref="GI:488729"
                     /db_xref="SWISS-PROT:Q29441"
                     /translation="MAKISTQYSHPTRTHPSVRTMDRDLDCIENGLSRTHLPCEETSS
                     ELQEGIAMETRGLAESRQSSFTSQGPTRLSRLIISLRAWSARHLHQEDQRPDSFLERF
                     RGAELQEVSSRESHVQFNVGSQEPPDRGRSAWPLARNNTNTCNNSEKDDKAKKEEKEK
                     KEEKKENPKKEEKKKDSVVMDPSSNMYYHWLTVIAVPVFYNWCLLVCRACFDELQSEH
                     LMLWLVLDYSADILYGMDMLVRARTGFLEQGLMVMDASRLWKHYTQTLHFKLDVLSLV
                     PTDLAYFKLGMNYPELRFNRLLKLARLFEFFDRTETRTNYPNMFRIGNLVLYILIIIH
                     WNACIYFAISKFIGFGTDSWVYPNVSNPEYGRLSRKYIYSLYWSTLTLTTIGETPPPV
                     KDEEYLFVVIDFLVGVLIFATIVGNVGSMISNMNASRAEFQAKIDSIKQYMQFRKVTK
                     DLETRVIRWFDYLWANKKTVDEKEVLKSLPDKLKAEIAINVHLDTLRKVRIFQDCEAG
                     LLVELVLKLRPAVFSPGDYICKKGDIGREMYIIKEGKLAVVAEDGITQFVVLGDGSYF
                     GEISILNIKGSKSGNRRTANIRSIGYSDLFCLSKDDLMEALTEYPEAKKALEEKGRQI
                     LMKDNLIDEELAKAGADPKDIEEKVEHLETSLDSLQTRFARLLAEYNATQMKVKQRLS
                     QLESQVKMGLPPDGDAPQTEASQP"
BASE COUNT      997 a   1054 c   1050 g    897 t
ORIGIN
```

```
alignment_scores:
            Quality: 1579.00              Length:    564
              Ratio:  3.486                Gaps:      3
 Percent Similarity:  80.319   Percent Identity: 52.660

alignment_block:
 US-09-927-267-1 x BTCNG3   ..

 Align seg 1/1  to: BTCNG3  from: 1  to: 3998

       3 GlnAspThrLysValLysThrThrGluSerSerPro..............  14
         :::|||   |||:::|||      |||::::::
     717 AAGGATGACAAGGCGAAAAAGGAGGAGAAAGAGAAAAAGGAAGAGAAAAA 766

      15 .......ProAlaProSerLysAlaArgLysLeuLeuProValLeuAspP 29
               |||      ::::::   :::|||         |||:::||||
     767 GGAGAACCCCAAGAAAGAGGAGAAGAAGAAGGACAGCGTTGTGATGGACC 816

      29 roSerGlyAspTyrTyrTyrTrpTrpLeuAsnThrMetValPheProVal 45
         |||||::::::   ||||||   ||||||:::::::::::::   ||||||
     817 CTTCCAGCAACATGTACTACCACTGGCTGACTGTCATCGCCGTGCCTGTC 866

      46 MetTyrAsnLeuIleIleLeuValCysArgAlaCysPheProAspLeuGl 62
         :::||||||      :::||||||||||||||||||||||   ::::|||||
     867 TTCTACAACTGGTGTCTGCTCGTGTGCAGGGCCTGTTTCGATGAGCTCCA 916

      62 nHisGlyTyrLeuValAlaTrpLeuValLeuAspTyrThrSerAspLeuL 79
         |      :::|||:::   |||||||||||||||||:::::|||:::|
     917 GTCCGAGCACCTGATGCTTTGGCTGGTCCTGGACTACTCAGCAGACATCC 966

      79 euTyrLeuLeuAspMetValValArgPheHisThrGlyPheLeuGluGln 95
         |||||   :::||||||:::||||||   :::||||||||||||||||||
     967 TCTATGGCATGGATATGCTGGTCCGAGCCCGGACAGGCTTCCTGGAGCAA 1016

      96 GlyIleLeuValValAspLysGlyArgIleSerSerArgTyrValArgTh 112
         |||:::::|||:::|||   :::|||:::   :::::::|||:::::||
    1017 GGCCTGATGGTCATGGACGCCAGCCGGCTGTGGAAGCACTACACACAGAC 1066

     112 rTrpSerPhePheLeuAspLeuAlaSerLeuMetProThrAspValValT 129
         |      |||   ||||||:::   ||||||:::|||||||||:::::::|
    1067 CTTGCACTTCAAGCTGGACGTGTTGTCCCTGGTGCCCACAGACCTGGCTT 1116

     129 yrValArgLeuGlyProHisThrProThrLeuArgLeuAsnArgPheLeu 145
         ||   :::||||||   :::   |||   ||||||:::|||||||:::|||
    1117 ATTTTAAGCTGGGCATGAACTACCCAGAACTGAGGTTCAACCGCCTCCTG 1166

     146 ArgAlaProArgLeuPheGluAlaPheAspArgThrGluThrArgThrAl 162
         :::      ||||||||||||   ||||||||||||||||||||||||
    1167 AAGTTGGCCCCGGCTCTTCGAATTCTTTGACCGCACGGAGACAAGGACCAA 1216

     162 aTyrProAsnAlaPheArgIleAlaLysLeuMetLeuTyrIlePheValV 179
          |||||||||   |||||||||:::::::|||:::|||||||||:::::::
    1217 CTACCCCAACATGTTCAGGATCGGGAACTTGGTCTTGTACATCCTCATCA 1266

     179 alIleHisTrpAsnSerCysLeuTyrPheAlaLeuSerArgTyrLeuGly 195
         ::|||||||||||||:::|||:::|||||||||:::|||:::::::::|||
    1267 TCATCCACTGGAATGCCTGCATCTACTTTGCCATTTCCAAGTTCATTGGT 1316

     196 PheGlyArgAspAlaTrpValTyrProAspProAlaGlnProGlyPheGl 212
         ||||||   |||:::|||||||||||:::   ::::::|||   :::
    1317 TTCGGGACCGACTCCTGGGTCTACCCAAACGTCTCCAACCCAGAGTATGG 1366

     212 uArgLeuArgArgGlnTyrLeuTyrSerPheTyrPheSerThrLeuIleL 229
          ||||||   |||:::|||:::||||||:::|||:::|||||||||   |
    1367 GCGCCTCTCCAGAAAGTACATTTACAGTCTCTACTGGTCCACCTTGACCC 1416

     229 euThrThrValGlyAspThrProProProAlaArgGluGluGluTyrLeu 245
         ||||||||:::|||:::|||||||||||||:::::::::|||||||||||
    1417 TGACCACCATTGGGGAGACCCCGCCCCCCGTAAAAGACGAGGAATATCTC 1466

     246 PheMetValGlyAspPheLeuLeuAlaValMetGlyPheAlaThrIleMe 262
         |||:::|||   |||||||||:::::::|||:::   ||||||||||||::
```

```
1467 TTTGTGGTCATCGACTTCCTGGTGGGCGTCCTGATTTTTGCCACCATCGT 1516

 262 tGlySerMetSerSerValIleTyrAsnMetAsnThrAlaAspAlaAlaP 279
     :|||::::::::::|||:::|||   |||||||||:::::::   |||   |
1517 GGGCAACGTGGGCTCCATGATCTCAAACATGAATGCTTCACGGGCCGAGT 1566

 279 heTyrProAspHisAlaLeuValLysLysTyrMetLysLeuGlnHisVal 295
     ||                  :::|||:::||||||:::::::::   |||
1567 TCCAGGCCAAGATCGATTCCATCAAGCAGTACATGCAGTTCCGCAAGGTG 1616

 296 AsnArgLysLeuGluArgArgValIleAspTrpTyrGlnHisLeuGlnIl 312
     ::::::   ||||||   |||||||||   |||::::::::::|||
1617 ACCAAGGACTTGGAGACACGGGTGATCCGCTGGTTCGACTACCTGTGGGC 1666

 312 eAsnLysLysMetThrAsnGluValAlaIleLeuGlnHisLeuProGluA 329
      |||||||||   :::::::|||      :::|||:::   ||||||::::
1667 CAATAAGAAGACAGTGGATGAGAAGGAGGTGCTCAAGAGCCTCCCCGACA 1716

 329 rgLeuArgAlaGluValAlaValSerValHisLeuSerThrLeuSerArg 345
     ::|||:::||||||:::|||:::::|||||||||:::||||||   :::
1717 AGCTGAAGGCCGAGATCGCCATCAACGTGCACCTGGACACCCTGAGGAAG 1766

 346 ValGlnIlePheGlnAsnCysGluAlaSerLeuLeuGluGluLeuValLe 362
     |||:::|||||||||:::|||||||||:::||||||   ||||||||||
1767 GTCCGAATCTTCCAGGACTGCGAGGCGGGGCTGCTGGTGGAGCTGGTGCT 1816

 362 uLysLeuGlnProGlnThrTyrSerProGlyGluTyrValCysArgLysG 379
     |||||||:::|||   :::::|||||||||:::|||:::|||:::||||
1817 GAAGCTGCGGCCGGCAGTGTTCAGCCCCGGGGACTACATCTGCAAGAAGG 1866

 379 lyAspIleGlyGlnGluMetTyrIleIleArgGluGlyGlnLeuAlaVal 395
     ||||||||||:::|||||||||||||||:::|||||:::||||||||||
1867 GGGACATCGGGAGGGAGATGTACATTATCAAGGAGGGCAAGCTGGCCGTG 1916

 396 ValAlaAspAspGlyIleThrGlnTyrAlaValLeuGlyAlaGlyLeuTy 412
     ||||||:::|||||||||||||||:::::|||||||||   |||   ||
1917 GTGGCCGAGGACGGCATCACCCAGTTCGTGGTCCTCGGCGACGGGAGTTA 1966

 412 rPheGlyGluIleSerIleIleAsnIleLysGlyAsnMetSerGlyAsnA 429
     ||||||||||||||||||:::|||||||||||||:::   ||||||||||
1967 TTTCGGGGAGATCAGCATCTTGAACATCAAGGGGAGCAAGTCCGGGAACC 2016

 429 rgArgThrAlaAsnIleLysSerLeuGlyTyrSerAspLeuPheCysLeu 445
     ||||||||||||||||||:::|||:::||||||||||||||||||||||
2017 GCCGCACGGCCAACATCAGGAGCATCGGCTACTCGGACCTGTTCTGCCTC 2066

 446 SerLysGluAspLeuArgGluValLeuSerGluTyrProGlnAlaGlnTh 462
     ||||||:::||||||   |||:::|||:::|||||||||:::|||:::
2067 TCCAAGGACGACCTGATGGAGGCGCTCACCGAGTACCCCGAGGCCAAGAA 2116

 462 rIleMetGluGluLysGlyArgGluIleLeuLeuLysMetAsnLysLeuA 479
       :::|||||||||||||||:::||||||:::|||   |||   :::|
2117 GGCGCTGGAGGAGAAAGGGCGGCAGATCCTCATGAAGGACAACCTGATCG 2166

 479 sp..............ValAsnAlaGluAlaAlaGluIleAlaLeuGln 490
     ||              ::::::|||:::      :::|||
2167 ACGAGGAGCTGGCCAAGGCCGGGGCAGACCCCAAGGACATC......... 2207

 491 GluAlaThrGluSerArgLeuArgGlyLeuAspGlnGlnLeuAspAspLe 507
             |||:::::::::::::   |||:::   :::|||||||:::||
2208 .........GAGGAGAAGGTGGAGCACCTCGAGACCTCCCTGGACTCCCT 2248

 507 uGlnThrLysPheAlaArgLeuLeuAlaGluLeuGluSerSerAlaLeuL 524
     |||||||:::|||||||||||||||||||||   :::::::::   :::|
2249 GCAGACCAGGTTTGCGCGGCTCCTGGCTGAGTACAACGCCACCCAGATGA 2298

 524 ysIleAlaTyrArgIleGluArgLeuGluTrpGlnThrArgGluTrpPro 540
     ||:::      |||:::::::::||||||   |||::::::
2299 AGGTGAAGCAGCGGCTCAGCCAGTTGGAAAGCCAGGTGAAGATGGGCCTC 2348

 541 MetProGluAspLeuAlaGluAlaAspAspGluGlyGluPro 554
        |||::      |||         :::   ::::::|||
2349 CCGCCTGATGGCGATGCTCCGCAAACTGAGGCCAGTCAGCCC 2390
```

```
seq_name: gb_ro:RNO272429

seq_documentation_block:
LOCUS       RNO272429              2098 bp    mRNA    linear   ROD 20-SEP-2000
DEFINITION  Rattus norvegicus mRNA for cyclic nucleotide-gated channel 2a,
            splice variant.
ACCESSION   AJ272429
VERSION     AJ272429.1  GI:10241816
KEYWORDS    cyclic nucleotide-gated channel.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 2098)
  AUTHORS   Meyer,M.R., Angele,A., Kremmer,E., Kaupp,U.B. and Muller,F.
  TITLE     A cGMP-signaling pathway in a subset of olfactory sensory neurons
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (19), 10595-10600 (2000)
   PUBMED   10984544
REFERENCE   2  (bases 1 to 2098)
  AUTHORS   Meyer,M.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-2000) Meyer M.R., Institut fuer Biologische
            Informationsverarbeitung - 1, Forschungszentrum Juelich,
            Leo-Brandt-Str., 52428 Juelich, GERMANY
FEATURES             Location/Qualifiers
     source          1. .2098
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /tissue_type="retina"
                     /tissue_type="olfactory epithelium"
     CDS             124. .2022
                     /note="splice variant"
                     /codon_start=1
                     /evidence=experimental
                     /product="cyclic nucleotide-gated channel 2a"
                     /protein_id="CAC09431.1"
                     /db_xref="GI:10241817"
                     /translation="MAKVNTQCSQPSPRQLSVKDAADRDLDQVENGLGRVSRFIISIR
                     AWAARHLHHEDPTPDSFLDRFHGAELKEVSSQERNAQPNPGGQEPPEGGKGRKKDPIV
                     VDPSSNIYYRWLTAIALPVFYNWCLLVCRACFDELQSEHLTLWLVLDYSADALYVVDM
                     LVRARTGFLEQGLMVRDTKRLWKHYTKTLHFKLDILSLIPTDLAYLKLGMNYPELRFN
                     RLLRFSRLFEFFDRTETRTNYPNVFRIGNLVLYTLIIIHWNACIYFAISKFIGFGTDS
                     WVYPNTSKPEYGRLSRKYIYSLYWSTLTLTTIGETPPPVKDEEYLFVVIDFLVGVLIF
                     ATIVGNVGSMISNMNASRAEFQAKIDSIKQYMQFRKVTKDLETRVIRWFDYLWANRKT
                     VDEKEVLKNLPDKLKAEIAINVHLDTLKKVRIFQDCEAGLLVELVLKLRPAVFSPGDY
                     ICKKGDIGREMYIIKEGKLAVVADDGVTQFVVLSDGSYFGEISILNIKGSKSGNRRTA
                     NIRSIGYSDLFCLSKDDLMETLTEYPDAKRALEEKGRQILMKDNLIDDDLVTARADAR
                     NIEEKVEYLESSLDGLQTRFARLLAEYSASQMKLKQRLSQLESQMTRRGHGFSPDREN
                     SEDASKAD"
BASE COUNT      510 a    551 c    590 g    447 t
ORIGIN

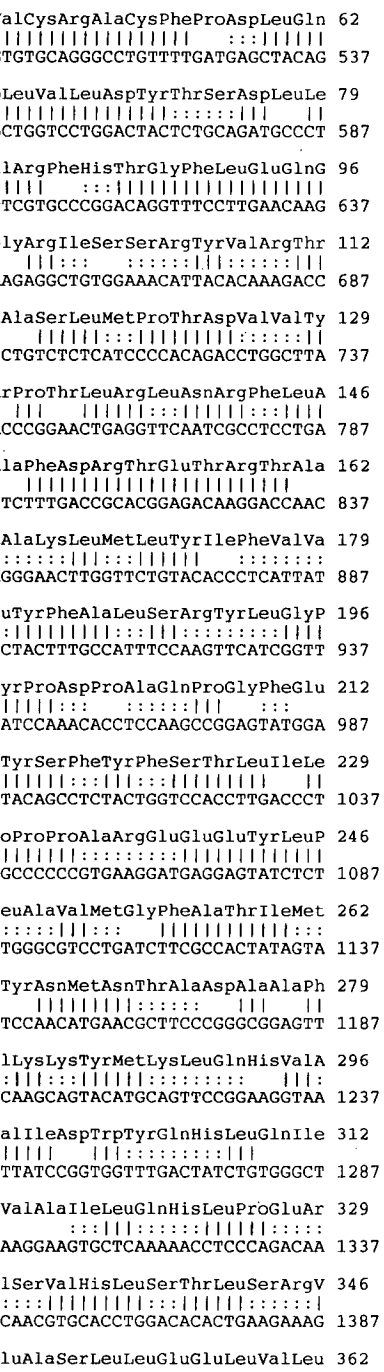

alignment_scores:
            Quality: 1578.00             Length:    550
              Ratio:  3.499               Gaps:      5
 Percent Similarity:  82.000   Percent Identity: 54.545

alignment_block:
 US-09-927-267-1 x RNO272429   ..

 Align seg 1/1  to: RNO272429  from: 1  to: 2098

      14 ProProAlaProSerLysAlaArgLysLeuLeuPro...ValLeuAspPr 29
         ||||||      :::|||:::||||||      |||   |||:::||||||
     388 CCACCAGAGGGAGGGAAAGGCAGGAAGAAGGATCCCATTGTGGTGGACCC 437

      29 oSerGlyAspTyrTyrTyrTrpTrpLeuAsnThrMetValProValM 46
         ||||::::::   ||||||   |||||||:::::::::::::|||||:
     438 CTCCAGCAACATCTACTACCGCTGGCTGACTGCCATCGCCCTCCCGGTCT 487
```



seq_name: gb_ro:RNO272428

seq_documentation_block:

LOCUS RNO272428 2212 bp mRNA linear ROD 20-SEP-2000
DEFINITION Rattus norvegicus mRNA for cyclic nucleotide-gated channel 2b.
ACCESSION AJ272428
VERSION AJ272428.1 GI:10241814
KEYWORDS cyclic nucleotide-gated channel.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 2212)
AUTHORS Meyer,M.R., Angele,A., Kremmer,E., Kaupp,U.B. and Muller,F.
TITLE A cGMP-signaling pathway in a subset of olfactory sensory neurons
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (19), 10595-10600 (2000)
PUBMED 10984544
REFERENCE 2 (bases 1 to 2212)
AUTHORS Meyer,M.R.
TITLE Direct Submission
JOURNAL Submitted (22-FEB-2000) Meyer M.R., Institut fuer Biologische

```
          Informationsverarbeitung - 1, Forschungszentrum Juelich,
          Leo-Brandt-Str., 52428 Juelich, GERMANY
FEATURES             Location/Qualifiers
     source          1. .2212
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /tissue_type="olfactory epithelium"
     CDS             124. .2136
                     /codon_start=1
                     /evidence=experimental
                     /product="cyclic nucleotide-gated channel 2b"
                     /protein_id="CAC09430.1"
                     /db_xref="GI:10241815"
                     /translation="MAKVNTQCSQPSPRQLSVKDAADRDLDQVENGLGRSHSPCEETS
                     STVHQGIVMETGELAGAAQSLFTSQGPARVSRFIISIRAWAARHLHHEDPTPDSFLDR
                     FHGAELKEVSSQERNAQPNPGGQEPPEGGKGRKKDPIVVDPSSNIYYRWLTAIALPVF
                     YNWCLLVCRACFDELQSEHLTLWLVLDYSADALYVVDMLVRARTGFLEQGLMVRDTKR
                     LWKHYTKTLHFKLDILSLIPTDLAYLKLGMNYPELRFNRLLRFSRLFEFFDRTETRTN
                     YPNVFRIGNLVLYTLIIIHWNACIYFAISKFIGFGTDSWVYPNTSKPEYGRLSRKYIY
                     SLYWSTLTLTTIGETPPPVKDEEYLFVVIDFLVGVLIFATIVGNVGSMISNMNASRAE
                     FQAKIDSIKQYMQFRKVTKDLETRVIRWFDYLWANRKTVDEKEVLKNLPDKLKAEIAI
                     NVHLDTLKKVRIFQDCEAGLLVELVLKLRPAVFSPGDYICKKGDIGREMYIIKEGKLA
                     VVADDGVTQFVVLSDGSYFGEISILNIKGSKSGNRRTANIRSIGYSDLFCLSKDDLME
                     TLTEYPDAKRALEEKGRQILMKDNLIDDDLVTARADARNIEEKVEYLESSLDGLQTRF
                     ARLLAEYSASQMKLKQRLSQLESQMTRRGHGFSPDRENSEDASKAD"
BASE COUNT      539 a    582 c    624 g    467 t
ORIGIN


alignment_scores:
            Quality: 1578.00              Length:    550
              Ratio:  3.499                Gaps:      5
 Percent Similarity:  82.000   Percent Identity: 54.545

alignment_block:
 US-09-927-267-1 x RNO272428   ..

 Align seg 1/1  to: RNO272428  from: 1  to: 2212

      14 ProProAlaProSerLysAlaArgLysLeuLeuPro...ValLeuAspPr 29
         ||||||      :::|||:::||||||      |||   |||:::|||||
     502 CCACCAGAGGGAGGGAAAGGCAGGAAGAAGGATCCCATTGTGGTGGACCC 551

      29 oSerGlyAspTyrTyrTyrTrpTrpLeuAsnThrMetValProProValM 46
         ||||::::::   ||||||   ||||||:::::::::::::::::||||||:
     552 CTCCAGCAACATCTACTACCGCTGGCTGACTGCCATCGCCCTCCCGGTCT 601

      46 etTyrAsnLeuIleIleLeuValCysArgAlaCysPheProAspLeuGln 62
         ::|||||||      :::|||||||||||||||||||||||   :::|||||||
     602 TCTATAACTGGTGTCTACTTGTGTGCAGGGCCTGTTTTGATGAGCTACAG 651

      63 HisGlyTyrLeuValAlaTrpLeuValLeuAspTyrThrSerAspLeuLe 79
            :::|||:::   ||||||||||||||||||:::::|||   ||
     652 TCAGAACACCTGACACTGTGGCTGGTCCTGGACTACTCTGCAGATGCCCT 701

      79 uTyrLeuLeuAspMetValValArgPheHisThrGlyPheLeuGluGlnG 96
         ||||::::::|||||:::||||||   :::|||||||||||||||||||
     702 GTATGTTGTGGACATGCTGGTTCGTGCCCGGACAGGTTTCCTTGAACAAG 751

      96 lyIleLeuValValAspLysGlyArgIleSerSerArgTyrValArgThr 112
         ||:::::::|||   |||      |||:::   ::::::|||:::::::|||
     752 GCCTAATGGTCAGGGATACCAAGAGGCTGTGGAAACATTACACAAAGACC 801

     113 TrpSerPhePheLeuAspLeuAlaSerLeuMetProThrAspValValTy 129
            |||   ||||||:::   ||||||:::||||||||:::::||
     802 TTGCACTTCAAGCTGGACATCCTGTCTCTCATCCCCACAGACCTGGCTTA 851

     129 rValArgLeuGlyProHisThrProThrLeuArgLeuAsnArgPheLeuA 146
         |:::::||||||   :::   |||   |||||:::||||||:::||||
     852 TTTGAAGTTGGGCATGAACTACCCGGAACTGAGGTTCAATCGCCTCCTGA 901

     146 rgAlaProArgLeuPheGluAlaPheAspArgThrGluThrArgThrAla 162
         ||      ||||||||||||   ||||||||||||||||||||||||
```

```
     902 GGTTCTCTCGGCTCTTTGAGTTCTTTGACCGCACGGAGACAAGGACCAAC 951

     163 TyrProAsnAlaPheArgIleAlaLysLeuMetLeuTyrIlePheValVa 179
         ||||||||||:::||||||||:::::::|||:::|||||   :::::::
     952 TACCCCAATGTGTTCAGGATAGGGAACTTGGTTCTGTACACCCTCATTAT 1001

     179 lIleHisTrpAsnSerCysLeuTyrPheAlaLeuSerArgTyrLeuGlyP 196
         :|||||||||||:::|||:::||||||||:::|||:::::::::||||
    1002 CATCCACTGGAACGCCTGCATCTACTTTGCCATTTCCAAGTTCATCGGTT 1051

     196 heGlyArgAspAlaTrpValTyrProAspProAlaGlnProGlyPheGlu 212
         |||||   |||:::||||||||||||:::   :::::::|||   :::
    1052 TTGGAACAGATTCCTGGGTCTATCCAAACACCTCCAAGCCGGAGTATGGA 1101

     213 ArgLeuArgArgGlnTyrLeuTyrSerPheTyrPheSerThrLeuIleLe 229
         ||||||   |||:::|||:::||||||:::|||:::||||||||   ||
    1102 CGCCTCTCCAGGAAGTACATTTACAGCCTCTACTGGTCCACCTTGACCCT 1151

     229 uThrThrValGlyAspThrProProProAlaArgGluGluGluTyrLeuP 246
         ||||||||:::|||:::|||||||||||||:::::::::||||||||||||
    1152 GACCACCATTGGGGAGACCCCGCCCCCCGTGAAGGATGAGGAGTATCTCT 1201

     246 heMetValGlyAspPheLeuLeuAlaValMetGlyPheAlaThrIleMet 262
         ||:::|||   |||||||||:::::|||:::   ||||||||||||:::
    1202 TTGTGGTCATAGACTTCCTGGTGGGCGTCCTGATCTTCGCCACTATAGTA 1251

     263 GlySerMetSerSerValIleTyrAsnMetAsnThrAlaAspAlaAlaPh 279
         |||:::::::::|||:::|||   ||||||||||:::::   |||   ||
    1252 GGCAATGTGGGCTCCATGATTTCCAACATGAACGCTTCCCGGGCGGAGTT 1301

     279 eTyrProAspHisAlaLeuValLysLysTyrMetLysLeuGlnHisValA 296
         |                  :::|||:::||||||:::::::   |||:
    1302 CCAGGCTAAGATAGATTCCATCAAGCAGTACATGCAGTTCCGGAAGGTAA 1351

     296 snArgLysLeuGluArgArgValIleAspTrpTyrGlnHisLeuGlnIle 312
         :::::   ||||||   |||||||||   |||:::::::::|||
    1352 CCAAGGACTTGGAGACTCGGGTTATCCGGTGGTTTGACTATCTGTGGGCT 1401

     313 AsnLysLysMetThrAsnGluValAlaIleLeuGlnHisLeuProGluAr 329
         |||:::|||   :::::::|||      :::|||:::::::||||||:::::
    1402 AACAGGAAGACGGTGGATGAAAAGGAAGTGCTCAAAAACCTCCCAGACAA 1451

     329 gLeuArgAlaGluValAlaValSerValHisLeuSerThrLeuSerArgV 346
         :|||:::||||||:::|||:::::|||||||||:::||||||:::::|
    1452 GCTGAAGGCCGAGATCGCCATCAACGTGCACCTGGACACACTGAAGAAAG 1501

     346 alGlnIlePheGlnAsnCysGluAlaSerLeuLeuGluGluLeuValLeu 362
         ||:::|||||||||:::||||||||:::||||||   ||||||||||||
    1502 TCCGAATCTTCCAGGACTGCGAGGCAGGGCTTCTGGTGGAGCTGGTGCTG 1551

     363 LysLeuGlnProGlnThrTyrSerProGlyGluTyrValCysArgLysGl 379
         ||||||:::|||   ::::::||||||||||:::|||:::|||:::||||||
    1552 AAGCTTCGTCCTGCTGTGTTCAGCCCTGGGGACTACATTTGCAAAAAGGG 1601

     379 yAspIleGlyGlnGluMetTyrIleIleArgGluGlyGlnLeuAlaValV 396
         ||||||||||:::|||||||||||||||:::|||||||:::||||||||||
    1602 GGACATTGGAAGGGAGATGTACATCATCAAGGAGGGCAAGCTGGCTGTCG 1651

     396 alAlaAspAspGlyIleThrGlnTyrAlaValLeuGlyAlaGlyLeuTyr 412
         ||||||||||||||:::||||||:::::::||||||:::   |||   |||
    1652 TGGCTGACGATGGGGTCACCCAGTTTGTGGTCCTCAGTGATGGCAGTTAC 1701

     413 PheGlyGluIleSerIleIleAsnIleLysGlyAsnMetSerGlyAsnAr 429
         ||||||||||||||||||:::||||||||||||:::   ||||||||||
    1702 TTTGGGGAGATCAGCATCTTAAACATTAAGGGGAGCAAGTCGGGGAACCG 1751

     429 gArgThrAlaAsnIleLysSerLeuGlyTyrSerAspLeuPheCysLeuS 446
         ||||||||||||||||:::|||:::||||||||||||||||||||||||||
    1752 CAGGACAGCCAACATCAGGAGCATCGGCTACTCGGACCTGTTCTGCCTCT 1801

     446 erLysGluAspLeuArgGluValLeuSerGluTyrProGlnAlaGlnThr 462
         |||||:::||||||   |||:::|||:::|||||||||:::|||:::
    1802 CCAAGGATGACTTGATGGAGACCCTCACGGAGTACCCGGATGCTAAGAGG 1851
```

```
     463 IleMetGluGluLysGlyArgGluIleLeuLeuLysMetAsnLysLeuAs 479
           :::|||||||||||||||||:::||||||:::|||   |||   :::||
    1852 GCTCTGGAGGAAAAGGGTCGGCAGATTCTGATGAAGGACAACCTAATCGA 1901

     479 p..............ValAsnAlaGluAlaAlaGluIleAlaLeuGlnG 491
         |              :::::::|||:::|||   :::|||
    1902 TGACGACCTGGTGACGGCCAGGGCAGATGCCAGGAACATC.......... 1941

     491 luAlaThrGluSerArgLeuArgGlyLeuAspGlnGlnLeuAspAspLeu 507
              |||::::::::::::   |||::::::::::||||||   |||
    1942 ........GAGGAGAAGGTGGAGTACCTGGAGTCATCCTTGGACGGCCTG 1983

     508 GlnThrLysPheAlaArgLeuLeuAlaGluLeuGluSerSerAlaLeuLy 524
         ||||||:::|||||||||||||||||||||||   :::::::|||   :::||
    1984 CAGACGAGGTTCGCCCGACTCCTGGCTGAGTACAGTGCCTCCCAGATGAA 2033

     524 sIleAlaTyrArgIleGluArgLeuGluTrpGln...ThrArg....... 537
         |:::      |||:::::::::||||||   |||   ||||||
    2034 GCTGAAACAGCGGCTCAGCCAGCTGGAGAGCCAGATGACCAGGAGGGGTC 2083

     538 ..............GluTrpProMetProGluAspLeuAlaGluAlaAsp 549
                       :::          ||||||   ::::::||||||
    2084 ATGGCTTCTCACCTGACAGGGAGAATTCCGAGGATGCTTCAAAAGCTGAC 2133

seq_name: gb_ro:AB002801

seq_documentation_block:
LOCUS       AB002801               2213 bp    mRNA    linear   ROD 14-JAN-1998
DEFINITION  Rattus norvegicus mRNA for cyclic nucleotide-gated channel,
            complete cds.
ACCESSION   AB002801
VERSION     AB002801.1  GI:2780733
KEYWORDS    cyclic nucleotide-gated channel.
SOURCE      Rattus norvegicus cDNA to mRNA.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 2213)
  AUTHORS   Misaka,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-APR-1997) Takumi Misaka, The University of Tokyo,
            Applied Biological Chemistry; 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113,
            Japan (E-mail:aa57035@hongo.ecc.u-tokyo.ac.jp,
            Tel:03-3812-2111(ex.5129), Fax:03-5802-8897)
REFERENCE   2  (sites)
  AUTHORS   Misaka,T., Kusakabe,Y., Emori,Y., Gonoi,T., Arai,S. and Abe,K.
  TITLE     Taste buds have a cyclic nucleotide-activated channel, CNGgust
  JOURNAL   J. Biol. Chem. 272 (36), 22623-22629 (1997)
  MEDLINE   97426406
FEATURES             Location/Qualifiers
     source          1. .2213
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
     CDS             14. .1849
                     /codon_start=1
                     /product="cyclic nucleotide-gated channel"
                     /protein_id="BAA24353.1"
                     /db_xref="GI:2780734"
                     /translation="MHQMETSTMVQDNRVSRFIISIRAWAARHLHHEDPTPDSFLDRF
                     HGAEPKEVSSQERNAQPNPGGQEPPEGGKGRKKDPIVVDPSSNIYYRWLTAIALPVFY
                     NWCLLVCRACFDELQSEHLTLWLVLDYSADALYVVDMLVRARTGFLEQGLMVRDTKRL
                     WKHYTKTLHFKLDILSLIPTDLAYLKLGMNYPELRFNRLLRFSRLFEFFDRTETRTNY
                     PNVFRIGNLVLYTLIIIHWNACIYFAISKFIGFGTDSWVYPNTSKPEYGRLSRKYIYS
                     LYWSTLTLTTIGETPPPVKDEEYLFVVIDFLVGVLIFATIVGNVGSMISNMNASRAEF
                     QAKIDSIKQYMQFRKVTKDLETRVIRWFDYLWANRKTVDEKEVLKNLPDKLKAEIAIN
                     VHLDTLKKVRIFQDCEAGLLVELVLKLRPAVFSPGDYICKKGDIGREMYIIKEGKLAV
                     VADDGVTQFVVLSDGSYFGEISILNIKGSKSGNRRTANIRSIGYSDLFCLSKDDLMET
                     LTEYPDAKRALEEKGRQILMKDNLIDDDLVTARADARNIEEKVEYLESSLDGLQTRFA
                     RLLAEYSASQMKLKQRLSQLESQMTRRGHGFSPDRENSEDASKAD"
BASE COUNT      535 a    552 c    616 g    510 t
ORIGIN
```

```
alignment_scores:
            Quality: 1578.00              Length:    550
              Ratio:  3.499                Gaps:      5
 Percent Similarity:  82.000   Percent Identity: 54.545

alignment_block:
 US-09-927-267-1 x AB002801   ..

 Align seg 1/1  to: AB002801  from: 1  to: 2213

      14 ProProAlaProSerLysAlaArgLysLeuLeuPro...ValLeuAspPr 29
         ||||||      :::|||:::||||||      |||   |||:::|||||
     215 CCACCAGAGGGAGGGAAAGGCAGGAAGAAGGATCCCATTGTGGTGGACCC 264

      29 oSerGlyAspTyrTyrTyrTrpTrpLeuAsnThrMetValProProValM 46
         ||||::::::   ||||||   ||||||:::::::::::::::||||||:
     265 CTCCAGCAACATCTACTACCGCTGGCTGACTGCCATCGCCCTCCCGGTCT 314

      46 etTyrAsnLeuIleIleLeuValCysArgAlaCysPheProAspLeuGln 62
         ::|||||      :::||||||||||||||||||||   :::||||||
     315 TCTATAACTGGTGTCTACTTGTGTGCAGGGCCTGTTTTGATGAGCTACAG 364

      63 HisGlyTyrLeuValAlaTrpLeuValLeuAspTyrThrSerAspLeuLe 79
               :::|||:::   ||||||||||||||||||:::::|||   ||
     365 TCAGAACACCTGACACTGTGGCTGGTCCTGGACTACTCTGCAGATGCCCT 414

      79 uTyrLeuLeuAspMetValValArgPheHisThrGlyPheLeuGluGlnG 96
         ||||:::::|||||:::||||||   :::|||||||||||||||||||||
     415 GTATGTTGTGGACATGCTGGTTCGTGCCCGGACAGGTTTCCTTGAACAAG 464

      96 lyIleLeuValValAspLysGlyArgIleSerSerArgTyrValArgThr 112
         ||:::::|||   |||      |||:::   ::::::|||:::::|||
     465 GCCTAATGGTCAGGGATACCAAGAGGCTGTGGAAACATTACACAAAGACC 514

     113 TrpSerPhePheLeuAspLeuAlaSerLeuMetProThrAspValValTy 129
             |||   ||||||:::   ||||||:::||||||||:::::::||
     515 TTGCACTTCAAGCTGGACATCCTGTCTCTCATCCCCACAGACCTGGCTTA 564

     129 rValArgLeuGlyProHisThrProThrLeuArgLeuAsnArgPheLeuA 146
         |:::::::||||||   :::   |||   ||||||:::||||||:::||||
     565 TTTGAAGTTGGGCATGAACTACCCGGAACTGAGGTTCAATCGCCTCCTGA 614

     146 rgAlaProArgLeuPheGluAlaPheAspArgThrGluThrArgThrAla 162
         ||      ||||||||||||   ||||||||||||||||||||||||
     615 GGTTCTCTCGGCTCTTTGAGTTCTTTGACCGCACGGAGACAAGGACCAAC 664

     163 TyrProAsnAlaPheArgIleAlaLysLeuMetLeuTyrIlePheValVa 179
         |||||||||:::||||||||:::::::|||:::||||||   ::::::::
     665 TACCCCAATGTGTTCAGGATAGGGAACTTGGTTCTGTACACCCTCATTAT 714

     179 lIleHisTrpAsnSerCysLeuTyrPheAlaLeuSerArgTyrLeuGlyP 196
         :|||||||||||||:::|||:::||||||||||:::|||:::::::::||||
     715 CATCCACTGGAACGCCTGCATCTACTTTGCCATTTCCAAGTTCATCGGTT 764

     196 heGlyArgAspAlaTrpValTyrProAspProAlaGlnProGlyPheGlu 212
         |||||   |||:::||||||||||||:::   :::::::|||   :::
     765 TTGGAACAGATTCCTGGGTCTATCCAAACACCTCCAAGCCGGAGTATGGA 814

     213 ArgLeuArgArgGlnTyrLeuTyrSerPheTyrPheSerThrLeuIleLe 229
         ||||||   |||:::|||:::||||||:::|||:::||||||||   ||
     815 CGCCTCTCCAGGAAGTACATTTACAGCCTCTACTGGTCCACCTTGACCCT 864

     229 uThrThrValGlyAspThrProProProAlaArgGluGluGluTyrLeuP 246
         |||||||:::|||:::||||||||||||:::::::::|||||||||||||
     865 GACCACCATTGGGGAGACCCCGCCCCCCGTGAAGGATGAGGAGTATCTCT 914

     246 heMetValGlyAspPheLeuLeuAlaValMetGlyPheAlaThrIleMet 262
         ||:::|||   |||||||||:::::|||:::   ||||||||||||:::
     915 TTGTGGTCATAGACTTCCTGGTGGGCGTCCTGATCTTCGCCACTATAGTA 964

     263 GlySerMetSerSerValIleTyrAsnMetAsnThrAlaAspAlaAlaPh 279
```

```
     |||:::::::::|||:::|||   |||||||||:::::   |||   ||
  965 GGCAATGTGGGCTCCATGATTTCCAACATGAACGCTTCCCGGGCGGAGTT 1014

  279 eTyrProAspHisAlaLeuValLysLysTyrMetLysLeuGlnHisValA 296
      |           :::|||:::|||||:::::::::   |||:
 1015 CCAGGCTAAGATAGATTCCATCAAGCAGTACATGCAGTTCCGGAAGGTAA 1064

  296 snArgLysLeuGluArgArgValIleAspTrpTyrGlnHisLeuGlnIle 312
      :::::   ||||||   |||||||||   |||::::::::::|||
 1065 CCAAGGACTTGGAGACTCGGGTTATCCGGTGGTTTGACTATCTGTGGGCT 1114

  313 AsnLysLysMetThrAsnGluValAlaIleLeuGlnHisLeuProGluAr 329
      |||:::|||   :::::::|||      :::|||:::::::||||||:::::
 1115 AACAGGAAGACGGTGGATGAAAAGGAAGTGCTCAAAAACCTCCCAGACAA 1164

  329 gLeuArgAlaGluValAlaValSerValHisLeuSerThrLeuSerArgV 346
      :|||:::||||||:::|||:::::::|||||||||:::||||||:::::|
 1165 GCTGAAGGCCGAGATCGCCATCAACGTGCACCTGGACACACTGAAGAAAG 1214

  346 alGlnIlePheGlnAsnCysGluAlaSerLeuLeuGluGluLeuValLeu 362
      ||:::|||||||||:::|||||||||:::||||||   ||||||||||||
 1215 TCCGAATCTTCCAGGACTGCGAGGCAGGGCTTCTGGTGGAGCTGGTGCTG 1264

  363 LysLeuGlnProGlnThrTyrSerProGlyGluTyrValCysArgLysGl 379
      ||||||:::|||   ::::::||||||||||:::|||:::|||:::||||||
 1265 AAGCTTCGTCCTGCTGTGTTCAGCCCTGGGGACTACATTTGCAAAAAGGG 1314

  379 yAspIleGlyGlnGluMetTyrIleIleArgGluGlyGlnLeuAlaValV 396
      ||||||||||:::|||||||||||||||:::||||||:::||||||||||
 1315 GGACATTGGAAGGGAGATGTACATCATCAAGGAGGGCAAGCTGGCTGTCG 1364

  396 alAlaAspAspGlyIleThrGlnTyrAlaValLeuGlyAlaGlyLeuTyr 412
      ||||||||||||||:::||||||:::::::||||||:::   |||   |||
 1365 TGGCTGACGATGGGGTCACCCAGTTTGTGGTCCTCAGTGATGGCAGTTAC 1414

  413 PheGlyGluIleSerIleIleAsnIleLysGlyAsnMetSerGlyAsnAr 429
      ||||||||||||||||||:::|||||||||||:::   ||||||||||
 1415 TTTGGGGAGATCAGCATCTTAAACATTAAGGGGAGCAAGTCGGGGAACCG 1464

  429 gArgThrAlaAsnIleLysSerLeuGlyTyrSerAspLeuPheCysLeuS 446
      ||||||||||||||||:::|||:::||||||||||||||||||||||||||
 1465 CAGGACAGCCAACATCAGGAGCATCGGCTACTCGGACCTGTTCTGCCTCT 1514

  446 erLysGluAspLeuArgGluValLeuSerGluTyrProGlnAlaGlnThr 462
      |||||:::||||||   |||:::|||:::|||||||||:::|||:::
 1515 CCAAGGATGACTTGATGGAGACCCTCACGGAGTACCCGGATGCTAAGAGG 1564

  463 IleMetGluGluLysGlyArgGluIleLeuLeuLysMetAsnLysLeuAs 479
          :::|||||||||||||||:::||||||:::|||   |||   :::||
 1565 GCTCTGGAGGAAAAGGGTCGGCAGATTCTGATGAAGGACAACCTAATCGA 1614

  479 p..............ValAsnAlaGluAlaAlaGluIleAlaLeuGlnG 491
      |              ::::::|||:::|||   :::|||
 1615 TGACGACCTGGTGACGGCCAGGGCAGATGCCAGGAACATC.......... 1654

  491 luAlaThrGluSerArgLeuArgGlyLeuAspGlnGlnLeuAspAspLeu 507
              |||:::::::::::::   |||:::::::::||||||   |||
 1655 ........GAGGAGAAGGTGGAGTACCTGGAGTCATCCTTGGACGGCCTG 1696

  508 GlnThrLysPheAlaArgLeuLeuAlaGluLeuGluSerSerAlaLeuLy 524
      ||||||:::|||||||||||||||||||||   :::::::|||   :::||
 1697 CAGACGAGGTTCGCCCGACTCCTGGCTGAGTACAGTGCCTCCCAGATGAA 1746

  524 sIleAlaTyrArgIleGluArgLeuGluTrpGln...ThrArg....... 537
      |:::      |||:::::::::||||||   |||   ||||||
 1747 GCTGAAACAGCGGCTCAGCCAGCTGGAGAGCCAGATGACCAGGAGGGGTC 1796

  538 ..............GluTrpProMetProGluAspLeuAlaGluAlaAsp 549
                    :::             ||||||   ::::::||||||
 1797 ATGGCTTCTCACCTGACAGGGAGAATTCCGAGGATGCTTCAAAAGCTGAC 1846

seq_name: gb_om:BTCAMPGC
```

```
seq_documentation_block:
LOCUS       BTCAMPGC               3166 bp    mRNA    linear   MAM 11-MAR-1993
DEFINITION  B.taurus mRNA for cAMP-gated channel.
ACCESSION   X55010
VERSION     X55010.1  GI:287743
KEYWORDS    cAMP-grated channel.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 3166)
  AUTHORS   Ludwig,J., Margalit,T., Eismann,E., Lancet,D. and Kaupp,U.B.
  TITLE     Primary structure of cAMP-gated channel from bovine olfactory
            epithelium
  JOURNAL   FEBS Lett. 270 (1-2), 24-29 (1990)
  MEDLINE   91032022
FEATURES             Location/Qualifiers
     source          1. .3166
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /cell_type="olfactory epithelium"
     CDS             418. .2409
                     /codon_start=1
                     /product="cAMP-gated channel"
                     /protein_id="CAA38754.1"
                     /db_xref="GI:287744"
                     /db_xref="SWISS-PROT:Q03041"
                     /translation="MTEKANGVKSSPANNHNHHAPPAIKASGKDDHRASSRPQSAAAD
                     DTSSELQQLAEMDAPQQRRGGFRRIARLVGVLREWAYRNFREEEPRPDSFLERFRGPE
                     LHTVTTQQGDGKGDKDGEGKGTKKKFELFVLDPAGDWYYRWLFLIALPVLYNWCLLVA
                     RACFSDLQKGYYIVWLVLDYVSDVVYIADLFIRLRTGFLEQGLLVKDTKKLRDNYIHT
                     MQFKLDVASIIPTDLIYFAVGIHNPEVRFNRLLHFARMFEFFDRTETRTSYPNIFRIS
                     NLILYILIIIHWNACIYYAISKSIGFGVDTWVYPNITDPEYGYLSREYIYCLYWSTLT
                     LTTIGETPPPVKDEEYLFVIFDFLIGVLIFATIVGNVGSMISNMNATRAEFQAKIDAV
                     KHYMQFRKVSKEMEAKVIRWFDYLWTNKKSVDEREVLKNLPAKLRAEIAINVHLSTLK
                     KVRIFQDCEAGLLVELVLKLRPQVFSPGDYICRKGDIGKEMYIIKEGKLAVVADDGVT
                     QYALLSAGSCFGEISILNIKGSKMGNRRTANIRSLGYSDLFCLSKDDLMEAVTEYPDA
                     KRVLEERGREILMKEGLLDENEVAASMEVDVQEKLEQLETNMDTLYTRFARLLAEYTG
                     AQQKLKQRITVLETKMKQNNEDDSLSDGMNSPEPPAEKP"
BASE COUNT      734 a    831 c    883 g    718 t
ORIGIN


alignment_scores:
            Quality: 1568.00              Length:    539
              Ratio:  3.548                Gaps:      2
 Percent Similarity:  82.004   Percent Identity: 54.174

alignment_block:
 US-09-927-267-1 x BTCAMPGC   ..

 Align seg 1/1  to: BTCAMPGC  from: 1  to: 3166

      18 SerLysAlaArgLysLeuLeuProValLeuAspProSerGlyAspTyrTy 34
         :::|||   :::      |||   |||||||||||:::||||||:::||
     787 ACCAAGAAGAAGTTTGAACTCTTTGTCTTGGACCCAGCCGGGGACTGGTA 836

      34 rTyrTrpTrpLeuAsnThrMetValPheProValMetTyrAsnLeuIleI 51
         ||||   ||||||      :::::::::||||||:::||||||      :
     837 CTACCGCTGGCTTTTTCTCATTGCCTTGCCCGTCCTCTACAACTGGTGCC 886

      51 leLeuValCysArgAlaCysPheProAspLeuGlnHisGlyTyrLeuVal 67
         ::||||||:::|||||||||||   |||||||||   ||||||   :::
     887 TATTGGTGGCCAGAGCCTGCTTCAGTGACCTGCAGAAAGGCTACTACATA 936

      68 AlaTrpLeuValLeuAspTyrThrSerAspLeuLeuTyrLeuLeuAspMe 84
         :::|||||||||||||||||:::||||||:::::||||:::   |||::
     937 GTGTGGCTGGTGCTGGATTACGTCTCAGATGTGGTCTACATCGCAGACCT 986

      84 tValValArgPheHisThrGlyPheLeuGluGlnGlyIleLeuValValA 101
         :   :::|||:::::::|||||||||||||||||||||:::||||||   |
     987 CTTCATCCGACTGCGCACAGGTTTCTTGGAGCAGGGGCTACTGGTGAAAG 1036
```

```
 101 spLysGlyArgIleSerSerArgTyrValArgThrTrpSerPhePheLeu 117
1037 ACACCAAGAAGTTGCGGGACAACTACATCCACACCATGCAGTTTAAGCTG 1086

 118 AspLeuAlaSerLeuMetProThrAspValValTyrValArgLeuGlyPr 134
1087 GATGTGGCCTCCATCATCCCTACAGACCTGATCTATTTTGCTGTGGGGAT 1136

 134 oHisThrProThrLeuArgLeuAsnArgPheLeuArgAlaProArgLeuP 151
1137 CCATAACCCTGAGGTGCGCTTCAACCGCCTGCTACACTTTGCCCGCATGT 1186

 151 heGluAlaPheAspArgThrGluThrArgThrAlaTyrProAsnAlaPhe 167
1187 TTGAGTTCTTTGACCGCACTGAGACACGCACCAGCTACCCCAACATCTTC 1236

 168 ArgIleAlaLysLeuMetLeuTyrIlePheValValIleHisTrpAsnSe 184
1237 CGAATAAGCAACCTGATCCTCTACATCTTGATCATCATTCACTGGAATGC 1286

 184 rCysLeuTyrPheAlaLeuSerArgTyrLeuGlyPheGlyArgAspAlaT 201
1287 CTGCATCTACTATGCCATCTCCAAGTCCATCGGCTTTGGGGTAGACACCT 1336

 201 rpValTyrProAspProAlaGlnProGlyPheGluArgLeuArgArgGln 217
1337 GGGTTTACCCCAACATCACTGACCCTGAGTATGGCTACCTGTCTAGGGAG 1386

 218 TyrLeuTyrSerPheTyrPheSerThrLeuIleLeuThrThrValGlyAs 234
1387 TACATCTATTGCCTTTACTGGTCTACACTGACCCTCACCACCATTGGGGA 1436

 234 pThrProProProAlaArgGluGluGluTyrLeuPheMetValGlyAspP 251
1437 GACACCACCCCCTGTAAAGGATGAGGAGTACCTGTTTGTCATCTTTGACT 1486

 251 heLeuLeuAlaValMetGlyPheAlaThrIleMetGlySerMetSerSer 267
1487 TCCTGATTGGTGTCCTCATCTTTGCCACCATCGTGGGAAATGTGGGCTCC 1536

 268 ValIleTyrAsnMetAsnThrAlaAspAlaAlaPheTyrProAspHisAl 284
1537 ATGATCTCCAACATGAATGCCACCCGGGCTGAGTTCCAGGCCAAGATTGA 1586

 284 aLeuValLysLysTyrMetLysLeuGlnHisValAsnArgLysLeuGluA 301
1587 TGCTGTCAAACATTATATGCAGTTCCGAAAGGTCAGCAAGGAGATGGAAG 1636

 301 rgArgValIleAspTrpTyrGlnHisLeuGlnIleAsnLysLysMetThr 317
1637 CCAAGGTCATTAGGTGGTTTGACTACTTGTGGACCAATAAGAAGAGTGTA 1686

 318 AsnGluValAlaIleLeuGlnHisLeuProGluArgLeuArgAlaGluVa 334
1687 GATGAGCGAGAAGTCCTCAAAAACCTGCCAGCAAAGCTCAGGGCTGAGAT 1736

 334 lAlaValSerValHisLeuSerThrLeuSerArgValGlnIlePheGlnA 351
1737 AGCCATCAACGTCCACCTGTCCACACTCAAGAAAGTGCGCATCTTTCAGG 1786

 351 snCysGluAlaSerLeuLeuGluGluLeuValLeuLysLeuGlnProGln 367
1787 ACTGTGAGGCTGGCCTGCTGGTGGAACTGGTATTAAAGCTCCGGCCTCAG 1836

 368 ThrTyrSerProGlyGluTyrValCysArgLysGlyAspIleGlyGlnGl 384
1837 GTCTTTAGCCCTGGGGACTACATTTGCCGCAAGGGGGATATTGGGAAGGA 1886

 384 uMetTyrIleIleArgGluGlyGlnLeuAlaValValAlaAspAspGlyI 401
1887 GATGTACATAATCAAGGAGGGAAAATTGGCAGTGGTGGCTGATGACGGTG 1936

 401 leThrGlnTyrAlaValLeuGlyAlaGlyLeuTyrPheGlyGluIleSer 417
1937 TCACTCAGTATGCCCTGCTCTCGGCTGGGAGTTGCTTTGGAGAGATCAGT 1986

 418 IleIleAsnIleLysGlyAsnMetSerGlyAsnArgArgThrAlaAsnIl 434
1987 ATCCTTAATATTAAGGGCAGCAAAATGGGCAATCGGCGCACAGCCAACAT 2036

 434 eLysSerLeuGlyTyrSerAspLeuPheCysLeuSerLysGluAspLeuA 451
2037 CCGCAGTCTTGGCTACTCTGATCTGTTCTGCTTGTCCAAGGATGATCTTA 2086

 451 rgGluValLeuSerGluTyrProGlnAlaGlnThrIleMetGluGluLys 467
2087 TGGAAGCTGTGACTGAGTACCCTGATGCCAAGAGGGTCTTGGAGGAGAGA 2136

 468 GlyArgGluIleLeuLeuLysMetAsnLysLeuAspValAsnAlaGluAl 484
2137 GGCCGGGAGATTCTGATGAAGGAGGGCTTGTTGGATGAGAATGAGGTGGC 2186

 484 aAla......GluIleAlaLeuGlnGluAlaThrGluSerArgLeuArgG 499
2187 AGCCAGCATGGAGGTAGATGTGCAGGAA.............AAGCTAGAAC 2224

 499 lyLeuAspGlnGlnLeuAspAspLeuGlnThrLysPheAlaArgLeuLeu 515
2225 AGCTGGAGACCAACATGGACACCTTGTACACTCGTTTTGCCCGCCTGCTG 2274

 516 AlaGluLeuGluSerSerAlaLeuLysIleAlaTyrArgIleGluArgLe 532
2275 GCCGAGTACACGGGAGCCCAGCAGAAGCTCAAGCAGCGCATCACAGTTTT 2324

 532 uGluTrpGlnThrArgGluTrpProMetProGluAspLeuAlaGluAlaA 549
2325 GGAAACGAAGATGAAGCAGAATAATGAGGATGACTCCCTGTCAGATGGGA 2374

 549 spAspGluGlyGluPro 554
2375 TGAACAGCCCAGAGCCA 2391
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2002, 14:35:16 ; Search time 7623.77 Seconds
(without alignments)
98.817 Million cell updates/sec

Title: US-09-927-267-12
Perfect score: 36
Sequence: 1 gcgaaagcttccaccatgagccaggacaccaaagtg 36

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	24	66.7	1835	6	AX350402	AX350402 Sequence
2	24	66.7	2013	6	AX355967	AX355967 Sequence
3	24	66.7	2551	6	AX350404	AX350404 Sequence
4	24	66.7	162965	2	AC021935	AC021935 Homo sapi
5	24	66.7	164434	2	AC036216	AC036216 Homo sapi
6	24	66.7	186609	2	AC022762	AC022762 Homo sapi
7	23.6	65.6	1828	10	MM4F2H	X14309 Murine mRNA
8	23.6	65.6	1838	10	AB023408	AB023408 Mus muscu
9	23.6	65.6	1861	6	AX305509	AX305509 Sequence
10	23.6	65.6	1861	10	MMU25708	U25708 Mus musculu
11	23.6	65.6	1902	10	U93712	U93712 Cricetulus
c 12	23.6	65.6	226399	2	AC025794	AC025794 Mus muscu
13	23.4	65.0	137154	9	AC019252	AC019252 Homo sapi
14	23.2	64.4	2732	9	HSU66464	U66464 Human hemat
15	23.2	64.4	126682	9	AC008649	AC008649 Homo sapi
16	23.2	64.4	191311	2	AC091486	AC091486 Homo sapi
17	22.4	62.2	1797	6	E41372	E41372 Amino acid
18	22.4	62.2	1797	10	RNU59324	U59324 Rattus norv
19	22.4	62.2	1907	10	RNRNAPLSL	X89225 R.norvegicu
20	22.4	62.2	1940	6	E58470	E58470 Neutral ami
21	22.4	62.2	1940	10	AB015433	AB015433 Rattus no
c 22	22.4	62.2	48883	2	AC095936	AC095936 Rattus no
23	22.4	62.2	81483	9	AC092182	AC092182 Homo sapi
c 24	22.4	62.2	181234	2	AC103148	AC103148 Rattus no
c 25	22.4	62.2	189481	2	AC026060	AC026060 Homo sapi
c 26	22.4	62.2	240021	2	AC099436	AC099436 Rattus no
27	22.4	62.2	283698	2	AC099294	AC099294 Rattus no
c 28	22.2	61.7	448	11	G53601	G53601 SHGC-86294
c 29	22.2	61.7	67702	9	AC104091	AC104091 Homo sapi
c 30	22.2	61.7	124015	9	AC006509	AC006509 Homo sapi
31	22.2	61.7	168395	2	AC092678	AC092678 Homo sapi
32	22	61.1	1822	9	HUM4F2A	J02769 Human 4F2 a
33	22	61.1	1861	9	HUMANTLA	J03569 Human lymph
34	22	61.1	1863	6	E41371	E41371 Amino acid
35	22	61.1	1863	9	HUMG4F2	J02939 Human membr
36	22	61.1	1879	9	AB018010	AB018010 Homo sapi
37	22	61.1	1897	6	E58472	E58472 Neutral ami
38	22	61.1	1964	6	AX281621	AX281621 Sequence
39	22	61.1	2304	9	HUM4F2HG1	M21898 Human 4F2 g
40	22	61.1	152883	2	AP000438	AP000438 Homo sapi
c 41	22	61.1	178404	2	AC090321	AC090321 Homo sapi
c 42	22	61.1	186326	2	AP001160	AP001160 Homo sapi
43	21.8	60.6	2020	9	BC003000	BC003000 Homo sapi
44	21.8	60.6	2188	9	BC001061	BC001061 Homo sapi
45	21.8	60.6	2247	9	AK025584	AK025584 Homo sapi

ALIGNMENTS

RESULT 1
AX350402
LOCUS AX350402 1835 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 27 from Patent WO0181578.
ACCESSION AX350402
VERSION AX350402.1 GI:18616045
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Vernet,C.A., Fernandes,E.R., Gerlach,V., Shimkets,R.A.,
Malyankar,U.M., Boldog,F.L., Zerhusen,B.D., Spytek,K.A.,
Majumder,K., Tchernev,V.T., Padigaru,M., Patturajan,M.,
Burgess,C.E., Gangolli,E.A., Smithson,G., Rastelli,L.,
Macdougall,J.R., Taupier,R.J., Grosse,W.M. and Alsobrook,J.P.
TITLE Novel proteins and nucleic acids encoding same
JOURNAL Patent: WO 0181578-A 27 01-NOV-2001;
Curagen Corporation (US)
FEATURES Location/Qualifiers
source 1. .1835
/organism="Homo sapiens"

```
                     /db_xref="taxon:9606"
BASE COUNT      420 a    521 c    521 g    373 t
ORIGIN


  Query Match             66.7%;  Score 24;  DB 6;  Length 1835;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 accatgagccaggacaccaaagtg 36
           ||||||||||||||||||||||||
Db      51 ACCATGAGCCAGGACACCAAAGTG 74


RESULT   2
AX355967
LOCUS       AX355967                2013 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 2 from Patent WO0202633.
ACCESSION   AX355967
VERSION     AX355967.1  GI:18620561
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Raumann,B.E., Sanjanwala,M.S., Tribouley,C.M. and Walia,N.K.
  TITLE     Transporters and ion channels
  JOURNAL   Patent: WO 0202633-A 2 10-JAN-2002;
            Incyte Genomics, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .2013
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /note="Incyte ID No: 6703242CB1"
BASE COUNT      472 a    576 c    557 g    408 t
ORIGIN


  Query Match             66.7%;  Score 24;  DB 6;  Length 2013;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 accatgagccaggacaccaaagtg 36
           ||||||||||||||||||||||||
Db     121 ACCATGAGCCAGGACACCAAAGTG 144


RESULT   3
AX350404
LOCUS       AX350404                2551 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 29 from Patent WO0181578.
ACCESSION   AX350404
VERSION     AX350404.1  GI:18616046
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Vernet,C.A., Fernandes,E.R., Gerlach,V., Shimkets,R.A.,
            Malyankar,U.M., Boldog,F.L., Zerhusen,B.D., Spytek,K.A.,
            Majumder,K., Tchernev,V.T., Padigaru,M., Patturajan,M.,
            Burgess,C.E., Gangolli,E.A., Smithson,G., Rastelli,L.,
            Macdougall,J.R., Taupier,R.J., Grosse,W.M. and Alsobrook,J.P.
  TITLE     Novel proteins and nucleic acids encoding same
  JOURNAL   Patent: WO 0181578-A 29 01-NOV-2001;
            Curagen Corporation (US)
FEATURES             Location/Qualifiers
     source          1. .2551
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      646 a    668 c    707 g    530 t
ORIGIN


  Query Match             66.7%;  Score 24;  DB 6;  Length 2551;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 accatgagccaggacaccaaagtg 36
           ||||||||||||||||||||||||
Db     776 ACCATGAGCCAGGACACCAAAGTG 799


RESULT   4
AC021935
LOCUS       AC021935              162965 bp    DNA     linear   HTG 03-MAR-2001
DEFINITION  Homo sapiens chromosome 11 clone RP11-645I8 map 11, WORKING DRAFT
            SEQUENCE, 5 unordered pieces.
ACCESSION   AC021935
VERSION     AC021935.3  GI:13184160
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 162965)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-645I8
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 162965)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 2, 2001 this sequence version replaced gi:7652890.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5277
                Center clone name: 645_I_8
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 40% of reads
                Sequencing vector: Plasmid; n/a; 60% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 161664 bases at least Q40
                Consensus quality: 161989 bases at least Q30
                Consensus quality: 162229 bases at least Q20
                Insert size: 158000; agarose-fp
                Insert size: 162565; sum-of-contigs
                Quality coverage: 12.2 in Q20 bases; agarose-fp
                Quality coverage: 11.9 in Q20.
```

```
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     5772: contig of 5772 bp in length
            *     5773 5872: gap of      100 bp
            *     5873   114225: contig of 108353 bp in length
            *   114226 114325: gap of      100 bp
            *   114326   119496: contig of 5171 bp in length
            *   119497 119596: gap of      100 bp
            *   119597   161781: contig of 42185 bp in length
            *   161782 161881: gap of      100 bp
            *   161882   162965: contig of 1084 bp in length.
FEATURES             Location/Qualifiers
     source          1. .162965
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-645I8"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .5772
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    5873. .114225
                     /note="assembly_fragment"
     misc_feature    114326. .119496
                     /note="assembly_fragment"
     misc_feature    119597. .161781
                     /note="assembly_fragment"
     misc_feature    161882. .162965
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    54235 a  33886 c  32984 g  41460 t    400 others
ORIGIN


  Query Match             66.7%;  Score 24;  DB 2;  Length 162965;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     13 accatgagccaggacaccaaagtg 36
          ||||||||||||||||||||||||
Db 149664 ACCATGAGCCAGGACACCAAAGTG 149687


RESULT   5
AC036216
LOCUS       AC036216               164434 bp    DNA     linear   HTG 04-MAY-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-658A5 map 11, WORKING DRAFT
            SEQUENCE, 26 unordered pieces.
ACCESSION   AC036216
VERSION     AC036216.2  GI:7705203
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 164434)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-658A5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 164434)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 4, 2000 this sequence version replaced gi:7523886.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L9404
                Center clone name: 658_A_5
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 149016 bases at least Q40
                Consensus quality: 157301 bases at least Q30
                Consensus quality: 160298 bases at least Q20
                Insert size: 157000; agarose-fp
                Insert size: 161934; sum-of-contigs
                Quality coverage: 3.7 in Q20 bases; agarose-fp
                Quality coverage: 3.6 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1416: contig of 1416 bp in length
            *     1417 1516: gap of      100 bp
            *     1517     3580: contig of 2064 bp in length
            *     3581 3680: gap of      100 bp
            *     3681     4383: contig of 703 bp in length
            *     4384 4483: gap of      100 bp
            *     4484     5844: contig of 1361 bp in length
            *     5845 5944: gap of      100 bp
            *     5945     7574: contig of 1630 bp in length
            *     7575 7674: gap of      100 bp
            *     7675     9988: contig of 2314 bp in length
            *     9989 10088: gap of      100 bp
            *    10089    13426: contig of 3338 bp in length
            *    13427 13526: gap of      100 bp
            *    13527    15573: contig of 2047 bp in length
            *    15574 15673: gap of      100 bp
            *    15674    18974: contig of 3301 bp in length
            *    18975 19074: gap of      100 bp
            *    19075    21809: contig of 2735 bp in length
            *    21810 21909: gap of      100 bp
            *    21910    25085: contig of 3176 bp in length
```

```
            *    25086 25185: gap of      100 bp
            *    25186    29425: contig of 4240 bp in length
            *    29426 29525: gap of      100 bp
            *    29526    34133: contig of 4608 bp in length
            *    34134 34233: gap of      100 bp
            *    34234    40649: contig of 6416 bp in length
            *    40650 40749: gap of      100 bp
            *    40750    45891: contig of 5142 bp in length
            *    45892 45991: gap of      100 bp
            *    45992    53180: contig of 7189 bp in length
            *    53181 53280: gap of      100 bp
            *    53281    62817: contig of 9537 bp in length
            *    62818 62917: gap of      100 bp
            *    62918    70247: contig of 7330 bp in length
            *    70248 70347: gap of      100 bp
            *    70348    78754: contig of 8407 bp in length
            *    78755 78854: gap of      100 bp
            *    78855    86840: contig of 7986 bp in length
            *    86841 86940: gap of      100 bp
            *    86941    98610: contig of 11670 bp in length
            *    98611 98710: gap of      100 bp
            *    98711   111069: contig of 12359 bp in length
            *   111070 111169: gap of      100 bp
            *   111170   122730: contig of 11561 bp in length
            *   122731 122830: gap of      100 bp
            *   122831   135722: contig of 12892 bp in length
            *   135723 135822: gap of      100 bp
            *   135823   150963: contig of 15141 bp in length
            *   150964 151063: gap of      100 bp
            *   151064   164434: contig of 13371 bp in length.
FEATURES             Location/Qualifiers
     source          1. .164434
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-658A5"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1416
                     /note="assembly_fragment"
     misc_feature    1517. .3580
                     /note="assembly_fragment"
     misc_feature    3681. .4383
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    4484. .5844
                     /note="assembly_fragment"
     misc_feature    5945. .7574
                     /note="assembly_fragment"
     misc_feature    7675. .9988
                     /note="assembly_fragment"
     misc_feature    10089. .13426
                     /note="assembly_fragment"
     misc_feature    13527. .15573
                     /note="assembly_fragment"
     misc_feature    15674. .18974
                     /note="assembly_fragment"
     misc_feature    19075. .21809
                     /note="assembly_fragment"
     misc_feature    21910. .25085
                     /note="assembly_fragment"
     misc_feature    25186. .29425
                     /note="assembly_fragment"
     misc_feature    29526. .34133
                     /note="assembly_fragment"
     misc_feature    34234. .40649
                     /note="assembly_fragment"
     misc_feature    40750. .45891
                     /note="assembly_fragment"
     misc_feature    45992. .53180
                     /note="assembly_fragment"
     misc_feature    53281. .62817
                     /note="assembly_fragment"
     misc_feature    62918. .70247
                     /note="assembly_fragment"
     misc_feature    70348. .78754
                     /note="assembly_fragment"
     misc_feature    78855. .86840
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    86941. .98610
                     /note="assembly_fragment"
     misc_feature    98711. .111069
                     /note="assembly_fragment"
     misc_feature    111170. .122730
                     /note="assembly_fragment"
     misc_feature    122831. .135722
                     /note="assembly_fragment"
     misc_feature    135823. .150963
                     /note="assembly_fragment"
     misc_feature    151064. .164434
                     /note="assembly_fragment"
BASE COUNT    50098 a  34031 c  34547 g  43240 t   2518 others
ORIGIN


  Query Match             66.7%;  Score 24;  DB 2;  Length 164434;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 accatgagccaggacaccaaagtg 36
           ||||||||||||||||||||||||
Db 156637 ACCATGAGCCAGGACACCAAAGTG 156660


RESULT   6
AC022762
LOCUS       AC022762              186609 bp    DNA     linear   HTG 09-FEB-2002
DEFINITION  Homo sapiens chromosome 11 clone RP11-290F24 map 11, *** SEQUENCING
            IN PROGRESS ***, 1 ordered pieces.
ACCESSION   AC022762
VERSION     AC022762.6  GI:18642872
KEYWORDS    HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 186609)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-290F24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 186609)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 9, 2002 this sequence version replaced gi:18483561.
            All repeats were identified using RepeatMasker:
```

```
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5842
                Center clone name: 290_F_24
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   186609: contig of 186609 bp in length.
FEATURES              Location/Qualifiers
     source           1. .186609
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="11"
                      /map="11"
                      /clone="RP11-290F24"
                      /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    60812 a  39184 c  37829 g  48784 t
ORIGIN


  Query Match             66.7%;  Score 24;  DB 2;  Length 186609;
  Best Local Similarity   100.0%;  Pred. No. 19;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     13 accatgagccaggacaccaaagtg 36
          ||||||||||||||||||||||||
Db 155443 ACCATGAGCCAGGACACCAAAGTG 155466


RESULT   7
MM4F2H
LOCUS       MM4F2H                  1828 bp    mRNA    linear   ROD 12-SEP-1993
DEFINITION  Murine mRNA for 4F2 antigen heavy chain.
ACCESSION   X14309
VERSION     X14309.1  GI:49749
KEYWORDS    4F2 antigen heavy chain; glycoprotein.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1828)
  AUTHORS   Leiden,J.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-FEB-1989) Leiden J. M., Howard Hughes Medical
            Institute, MSRB-I Room 4510C, University of Michigan Medical
            Centre, Ann Arbor MI
REFERENCE   2  (bases 1 to 1828)
  AUTHORS   Parmacek,M.S., Karpinski,B.A., Gottesdiener,K.M., Thompson,C.B. and
            Leiden,J.M.
  TITLE     Structure, expression and regulation of the murine 4F2 heavy chain
  JOURNAL   Nucleic Acids Res. 17 (5), 1915-1931 (1989)
  MEDLINE   89183602
COMMENT     Data kindly reviewed (28-April-1989) by Leiden J.M.
FEATURES              Location/Qualifiers
     source           1. .1828
                      /organism="Mus musculus"
                      /strain="C57Bl/6xDBA/2 and ICR"
                      /db_xref="taxon:10090"
                      /clone="M4F2HC-27, M4F2HC-50"
                      /clone_lib="macrophage lambda gt11, pre B-cell lambda
                      gt10"
     CDS              93. .1673
                      /note="4F2 heavy chain (AA 1-526)"
                      /codon_start=1
                      /protein_id="CAA32490.1"
                      /db_xref="GI:49750"
                      /db_xref="MGD:MGI:96955"
                      /db_xref="SWISS-PROT:P10852"
                      /translation="MSQDTEVDMKDVELNELEPEKQPMNAADGAAAGEKNGLVKIKVA
                      EDETEAGVKFTGLSKEELLKVAGSPGWVRTRWALLLLFWLGWLGMLAGAVVIIVRAPR
                      CRELPVQRWWHKGALYRIGDLQAFVGRDAGGIAGLKSHLEYLSTLKVKGLVLGPIHKN
                      QKDEINETDLKQINPTLGSQEDFKDLLQSAKKKSIHIILDLTPNYQGQNAWFLPAQAD
                      IVATKMKEALSSWLQDGVDGFQFRDVGKLMNAPLYLAEWQNITKNLSEDRLLIAGTES
                      SDLQQIVNILESTSDLLLTSSYLSNSTFTGERTESLVTRFLNATGSQWCSWSVSQAGL
                      LADFIPDHLLRLYQLLLFTLPGTPVFSYGDELGLQGALPGQPAKAPLMPWNESSIFHI
                      PRPVSLNMTVKGQNEDPGSLLTQFRRLSDLRGKERSLLHGDFHALSSSPDLFSYIRHW
                      DQNERYLVVLNFRDSGRSARLGASNLPAGISLPASAKLLLSTDSARQSREEDTSLKLE
                      NLSLNPYEGLLLQFPFVA"
BASE COUNT      402 a    507 c    500 g    419 t
ORIGIN


  Query Match             65.6%;  Score 23.6;  DB 10;  Length 1828;
  Best Local Similarity   86.7%;  Pred. No. 51;
  Matches   26;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      7 gcttccaccatgagccaggacaccaaagtg 36
          ||   |||||||||||||||||| |||||
Db     84 GCAGGCACCATGAGCCAGGACACCGAAGTG 113


RESULT   8
AB023408
LOCUS       AB023408                1838 bp    mRNA    linear   ROD 05-FEB-2000
DEFINITION  Mus musculus mRNA for 4F2 heavy chain, complete cds.
ACCESSION   AB023408
VERSION     AB023408.1  GI:6906722
KEYWORDS    4F2 heavy chain; transmembrane protein.
SOURCE      Mus musculus (strain:ICR) adult cDNA to mRNA, clone_lib:mouse brain
            cDNA clone:m4F2hc.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Kanai,Y., Watanabe,M. and Endou,H.
  TITLE     Localization of expression of system L neutral amino acid
            transporter LAT1 in brain
  JOURNAL   Unpublished (1999)
REFERENCE   2  (bases 1 to 1838)
  AUTHORS   Kanai,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-FEB-1999) Yoshikatsu Kanai, Kyorin University School
            of Medicine, Department of Pharmacolgy and Toxicology; 6-20-2
            Shinkawa, Mitaka, Tokyo 181-8611, Japan
            (E-mail:QZV07422@nifty.ne.jp, Tel:81-422-47-5511(ex.3453),
            Fax:81-422-79-1321)
FEATURES              Location/Qualifiers
     source           1. .1838
                      /organism="Mus musculus"
                      /strain="ICR"
                      /db_xref="taxon:10090"
                      /clone="m4F2hc"
                      /clone_lib="mouse brain cDNA"
                      /dev_stage="adult"
     CDS              106. .1686
                      /function="cell surface antigen, amino acid transporter"
                      /note="m4F2hc"
                      /codon_start=1
                      /product="4F2 heavy chain"
                      /protein_id="BAA90555.1"
                      /db_xref="GI:6906723"
                      /translation="MSQDTEVDMKDVELNELEPEKQPMNAADGAAAGEKNGLVKIKVA
                      EDETEAGVKFTGLSKEELLKVAGSPGWVRTRWALLLLFWLGWLGMLAGAVVIIVRAPR
```

```
                     CRELPVQRWWHKGALYRIGDLQAFVGRDAGGIAGLKSHLEYLSTLKVKGLVLGPIHKN
                     QKDEINETDLKQINPTLGSQEDFKDLLQSAKKKSIHIILDLTPNYQGQNAWFLPAQAD
                     IVATKMKEALSSWLQDGVDGFQFRDVGKLMNAPLYLAEWQNITKNLSEDRLLIAGTES
                     SDLQQIVNILESTSDLLLTSSYLSNSTFTGERTESLVTRFLNATGSQWCSWSVSQAGL
                     LADFIPDHLLRLYQLLLFTLPGTPVFSYGDELGLQGALPGQPAKAPLMPWNESSIFHI
                     PRPVSLNMTVKGQNEDPGSLLTQFRRLSDLRGKERSLLHGDFHALSSSPDLFSYIRHW
                     DQNERYLVVLNFRDSGRSARLGASNLPAGISLPASAKLLLSTDSARQSREEDTSLKLE
                     NLSLNPYEGLLLQFPFVA"
     polyA_site      1838
                     /note="14 A nucleotides"
BASE COUNT      403 a    510 c    503 g    422 t
ORIGIN


  Query Match             65.6%;  Score 23.6;  DB 10;  Length 1838;
  Best Local Similarity   86.7%;  Pred. No. 51;
  Matches   26;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      7 gcttccaccatgagccaggacaccaaagtg 36
          ||   |||||||||||||||||| |||||
Db     97 GCAGGCACCATGAGCCAGGACACCGAAGTG 126


RESULT  9
AX305509
LOCUS       AX305509                1861 bp    DNA     linear   PAT 11-DEC-2001
DEFINITION  Sequence 260 from Patent WO0188188.
ACCESSION   AX305509
VERSION     AX305509.1  GI:17645008
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Ishikawa,K., Asai,S., Takahashi,Y., Nagata,T. and Ishii,Y.
  TITLE     Method for examining ischemic conditions
  JOURNAL   Patent: WO 0188188-A 260 22-NOV-2001;
            School Juridical Person Nihon University (JP)
FEATURES             Location/Qualifiers
     source          1. .1861
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
BASE COUNT      404 a    519 c    520 g    418 t
ORIGIN


  Query Match             65.6%;  Score 23.6;  DB 6;  Length 1861;
  Best Local Similarity   86.7%;  Pred. No. 51;
  Matches   26;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      7 gcttccaccatgagccaggacaccaaagtg 36
          ||   |||||||||||||||||| |||||
Db    101 GCAGGCACCATGAGCCAGGACACCGAAGTG 130


RESULT  10
MMU25708
LOCUS       MMU25708                1861 bp    mRNA    linear   ROD 13-JUL-1996
DEFINITION  Mus musculus CD98 heavy chain mRNA, complete cds.
ACCESSION   U25708
VERSION     U25708.1  GI:973181
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1861)
  AUTHORS   Warren,A.P., Patel,K., McConkey,D.J. and Palacios,R.
  TITLE     CD98: a type II transmembrane glycoprotein expressed from the
            beginning of primitive and definitive hematopoiesis may play a
            critical role in the development of hematopoietic cells
  JOURNAL   Blood 87 (9), 3676-3687 (1996)
  MEDLINE   96199404
REFERENCE   2  (bases 1 to 1861)
  AUTHORS   Warren,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-APR-1995) Ronald Palacios, Immunology, University of
            Texas MD Anderson Cancer Center, 1515 Holcombe Boulevard, Houston,
            TX 77030, USA
FEATURES             Location/Qualifiers
     source          1. .1861
                     /organism="Mus musculus"
                     /strain="B10.BR"
                     /db_xref="taxon:10090"
                     /clone="Joro 177-87"
                     /cell_line="FTH5"
                     /cell_type="lymphocyte"
     CDS             110. .1708
                     /function="hematopoietic cell development"
                     /note="type II transmembrane protein"
                     /codon_start=1
                     /product="CD98 heavy chain"
                     /protein_id="AAB03769.1"
                     /db_xref="GI:973182"
                     /translation="MSQDTEVDMKDVELNELEPEKQPMNAADGAAMSLAGAEKNGLVK
                     IKVAEDETEAGVAAKFTGLSKEELLKVAGSPGWVRTRWALLLLFWLGWLGMLAGAVVI
                     IVRAPRCRELPVQRWWHKGALYRIGDLQAFVGRDAGGIAGLKSHLEYLSTLKVKGLVL
                     GPIHKNQKDEINETDLKQINPTLGSQEDFKDLLQSAKKKSIHIILDLTPNYQGQNAWF
                     LPAQADIVATKMKEALSSWLQDGVDGFQFRDVGKLMNAPLYLAEWQNITKNLSEDRLV
                     IAGTESSDLQQIVNILDSTSDLLLTSSYLSNSTHTGERTESLDTRLVSATGSHWCSWS
                     VSQAALLADFIPDHLLRLYQLLLFTLPGTPVFSYGDELGLQGALPGQPAKAPLMPWNE
                     SSIFHIPRPVSLNMTVKGQNEDPGSLLTQFRRLSDLRGKERSLLHGDFHALSSTPDLF
                     SYIRHWDQNERYLVVLNFRDSGRSARLGASNLPAGISLPASAKLLLSTDSARQSREED
                     TSLKLENLSLNPYEGLLLQFPFVA"
BASE COUNT      404 a    519 c    520 g    418 t
ORIGIN


  Query Match             65.6%;  Score 23.6;  DB 10;  Length 1861;
  Best Local Similarity   86.7%;  Pred. No. 51;
  Matches   26;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      7 gcttccaccatgagccaggacaccaaagtg 36
          ||   |||||||||||||||||| |||||
Db    101 GCAGGCACCATGAGCCAGGACACCGAAGTG 130


RESULT  11
U93712
LOCUS       U93712                  1902 bp    mRNA    linear   ROD 23-JAN-1998
DEFINITION  Cricetulus griseus putative CD98 protein mRNA, complete cds.
ACCESSION   U93712
VERSION     U93712.1  GI:2804625
KEYWORDS    .
SOURCE      Chinese hamster.
  ORGANISM  Cricetulus griseus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
            Cricetulus.
REFERENCE   1  (bases 1 to 1902)
  AUTHORS   Fenczik,C.A., Sethi,T., Ramos,J.W., Hughes,P.E. and Ginsberg,M.H.
  TITLE     Complementation of dominant suppression implicates CD98 in integrin
            activation
  JOURNAL   Nature 390 (6655), 81-85 (1997)
  MEDLINE   98028524
REFERENCE   2  (bases 1 to 1902)
  AUTHORS   Fenczik,C.A. and Ginsberg,M.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAR-1997) Vascular Biology, Scripps Research
            Institute, 10550 N. Torrey Pines, La Jolla, CA 92037, USA
FEATURES             Location/Qualifiers
     source          1. .1902
                     /organism="Cricetulus griseus"
                     /db_xref="taxon:10029"
                     /cell_line="CHO"
```

```
     CDS             134. .1735
                     /codon_start=1
                     /product="putative CD98 protein"
                     /protein_id="AAB97631.1"
                     /db_xref="GI:2804626"
                     /translation="MSQDTEVDMKEVELNELEPEKQPMNAAAGAAAGEKNGLVKIKVA
                     EDEAEAEAGAKFTGLSKEELLKVAGSPGWVRTRWALLLLFWLGWLGMLAGAVVIIVRA
                     PRCRELPVQRWWHEGALYRVGDLQAFVGPNAGGIAGLKKHLEYLSTLKVKGLVLGPIH
                     KNQKDDVNGTDLKQIDPSLGSQEDFKDLLQSAKKKSIHIILDLTPNYRGQSAWFLPAQ
                     ADWVATKMKEALSSWLQEGVDGFQFRDVEKLMDASLYLAEWQNITKNFSEDRLLIAGT
                     DSSDLQQIVTILESTSDLLLTSSYLSNSSFTGEHTKLLVTRYLNATGSHWCSWSVSQA
                     GLLTASVPAQLLRLYQLMLFTLPGTPVFSYGDEIGLQEAVVPGQPVEAPFMPWDESSL
                     PQTPGHISTNMTVKGQNEDPGSLLTQFRWLSDLRGKERGKERSLLHGDFRALTSSPGI
                     FSYVRYWDQNERYLVVLNFQDVVLSARLEAANLPAGISLPASANLLFSTNSTRLGQEQ
                     GTSLSLENLYLNPHEGLLLHFPYVA"
BASE COUNT      406 a    517 c    536 g    443 t
ORIGIN


  Query Match             65.6%;  Score 23.6;  DB 10;  Length 1902;
  Best Local Similarity   86.7%;  Pred. No. 51;
  Matches   26;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       7 gcttccaccatgagccaggacaccaaagtg 36
           ||   |||||||||||||||||| |||||
Db     125 GCAGGCACCATGAGCCAGGACACCGAAGTG 154


RESULT  12
AC025794/c
LOCUS       AC025794               226399 bp    DNA     linear   HTG 13-JUN-2001
DEFINITION  Mus musculus clone RP23-62O7, WORKING DRAFT SEQUENCE, 25 unordered
            pieces.
ACCESSION   AC025794
VERSION     AC025794.4  GI:14389364
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 226399)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-62O7
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 226399)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 13, 2001 this sequence version replaced gi:8225449.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html

            ---------- Genome Center
               Center: Whitehead Institute/ MIT Center for Genome Research
               Center code: WIBR
               Web site: http://www-seq.wi.mit.edu
               Contact: sequence_submissions@genome.wi.mit.edu
            ---------- Project Information
               Center project name: L8057
               Center clone name: 62_O_7
            ---------- Summary Statistics
               Sequencing vector: M13; M77815; 45% of reads
               Sequencing vector: Plasmid; n/a; 55% of reads
               Chemistry: Dye-terminator Big Dye; 100% of reads
               Assembly program: Phrap; version 0.960731
               Consensus quality: 217695 bases at least Q40
               Consensus quality: 221726 bases at least Q30
               Consensus quality: 223237 bases at least Q20
               Insert size: 219000; agarose-fp
               Insert size: 223999; sum-of-contigs
               Quality coverage: 8.2 in Q20 bases; agarose-fp
               Quality coverage: 8.0 in Q20 bas.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 25 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    15086: contig of 15086 bp in length
            *    15087 15186: gap of      100 bp
            *    15187    15923: contig of 737 bp in length
            *    15924 16023: gap of      100 bp
            *    16024    16694: contig of 671 bp in length
            *    16695 16794: gap of      100 bp
            *    16795    17880: contig of 1086 bp in length
            *    17881 17980: gap of      100 bp
            *    17981    18579: contig of 599 bp in length
            *    18580 18679: gap of      100 bp
            *    18680    19268: contig of 589 bp in length
            *    19269 19368: gap of      100 bp
            *    19369    20014: contig of 646 bp in length
            *    20015 20114: gap of      100 bp
            *    20115    20821: contig of 707 bp in length
            *    20822 20921: gap of      100 bp
            *    20922    21533: contig of 612 bp in length
            *    21534 21633: gap of      100 bp
            *    21634    22538: contig of 905 bp in length
            *    22539 22638: gap of      100 bp
            *    22639    23338: contig of 700 bp in length
            *    23339 23438: gap of      100 bp
            *    23439    24125: contig of 687 bp in length
            *    24126 24225: gap of      100 bp
            *    24226    24941: contig of 716 bp in length
            *    24942 25041: gap of      100 bp
            *    25042    25887: contig of 846 bp in length
            *    25888 25987: gap of      100 bp
            *    25988    26876: contig of 889 bp in length
            *    26877 26976: gap of      100 bp
            *    26977    28437: contig of 1461 bp in length
            *    28438 28537: gap of      100 bp
            *    28538    30287: contig of 1750 bp in length
            *    30288 30387: gap of      100 bp
            *    30388    32661: contig of 2274 bp in length
            *    32662 32761: gap of      100 bp
            *    32762    36782: contig of 4021 bp in length
            *    36783 36882: gap of      100 bp
            *    36883    45293: contig of 8411 bp in length
            *    45294 45393: gap of      100 bp
            *    45394    61819: contig of 16426 bp in length
            *    61820 61919: gap of      100 bp
            *    61920   126287: contig of 64368 bp in length
            *   126288 126387: gap of      100 bp
            *   126388   149055: contig of 22668 bp in length
```

```
          *   149056 149155: gap of      100 bp
          *   149156   211213: contig of 62058 bp in length
          *   211214 211313: gap of      100 bp
          *   211314   226399: contig of 15086 bp in length.
FEATURES             Location/Qualifiers
     source          1. .226399
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP23-6207"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1. .15086
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    15187. .15923
                     /note="assembly_fragment"
     misc_feature    16024. .16694
                     /note="assembly_fragment"
     misc_feature    16795. .17880
                     /note="assembly_fragment"
     misc_feature    17981. .18579
                     /note="assembly_fragment"
     misc_feature    18680. .19268
                     /note="assembly_fragment"
     misc_feature    19369. .20014
                     /note="assembly_fragment"
     misc_feature    20115. .20821
                     /note="assembly_fragment"
     misc_feature    20922. .21533
                     /note="assembly_fragment"
     misc_feature    21634. .22538
                     /note="assembly_fragment"
     misc_feature    22639. .23338
                     /note="assembly_fragment"
     misc_feature    23439. .24125
                     /note="assembly_fragment"
     misc_feature    24226. .24941
                     /note="assembly_fragment"
     misc_feature    25042. .25887
                     /note="assembly_fragment"
     misc_feature    25988. .26876
                     /note="assembly_fragment"
     misc_feature    26977. .28437
                     /note="assembly_fragment"
     misc_feature    28538. .30287
                     /note="assembly_fragment"
     misc_feature    30388. .32661
                     /note="assembly_fragment"
     misc_feature    32762. .36782
                     /note="assembly_fragment"
     misc_feature    36883. .45293
                     /note="assembly_fragment"
     misc_feature    45394. .61819
                     /note="assembly_fragment"
     misc_feature    61920. .126287
                     /note="assembly_fragment"
     misc_feature    126388. .149055
                     /note="assembly_fragment"
     misc_feature    149156. .211213
                     /note="assembly_fragment"
     misc_feature    211314. .226399
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    61164 a  51232 c  50642 g  60960 t   2401 others
ORIGIN


  Query Match           65.6%;  Score 23.6;  DB 2;  Length 226399;
  Best Local Similarity  86.7%;  Pred. No. 27;
  Matches   26;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       7 gcttccaccatgagccaggacaccaaagtg 36
           ||   |||||||||||||||||| |||||
Db 143283 GCAGGCACCATGAGCCAGGACACCGAAGTG 143254


RESULT 13
AC019252
LOCUS       AC019252              137154 bp    DNA     linear   PRI 09-JAN-2002
DEFINITION  Homo sapiens chromosome 8, clone RP11-390M9, complete sequence.
ACCESSION   AC019252
VERSION     AC019252.9  GI:18093026
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 137154)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-390M9
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 137154)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-DEC-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 137154)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 137154)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
```

```
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jan 9, 2002 this sequence version replaced gi:15055325.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L2097
                Center clone name: 390_M_9
            --------------

            overlaps at SP6 end with L6131 [in finishing] which includes gap
            [same gap in L6131]. This overlap is being finished in L6131.
FEATURES             Location/Qualifiers
     source          1. .137154
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-390M9"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   1. .35
                     /rpt_family="(CA)n"
     repeat_region   complement(1514. .1690)
                     /rpt_family="AluSg/x"
     repeat_region   1695. .1730
                     /rpt_family="(TTTA)n"
     repeat_region   complement(1731. .3121)
                     /rpt_family="L1PA5"
     repeat_region   complement(3124. .5301)
                     /rpt_family="L1PA5"
     repeat_region   5298. .6096
                     /rpt_family="L1P"
     repeat_region   complement(7108. .7275)
                     /rpt_family="MLT1A2"
     repeat_region   7276. .7629
                     /rpt_family="THE1C"
     repeat_region   complement(7630. .7693)
                     /rpt_family="MLT1A2"
     repeat_region   8270. .8297
                     /rpt_family="AT_rich"
     repeat_region   complement(8439. .8734)
                     /rpt_family="AluY"
     repeat_region   8992. .9140
                     /rpt_family="L1ME"
     repeat_region   9699. .9732
                     /rpt_family="AT_rich"
     repeat_region   complement(10460. .10756)
                     /rpt_family="AluSc"
     repeat_region   complement(10841. .11556)
                     /rpt_family="L1MC1"
     repeat_region   11557. .11584
                     /rpt_family="(TTTA)n"
     repeat_region   complement(11585. .12399)
                     /rpt_family="L1MC1"
     repeat_region   complement(12400. .12702)
                     /rpt_family="AluY"
     repeat_region   complement(12703. .13500)
                     /rpt_family="L1MC1"
     repeat_region   complement(13851. .14074)
                     /rpt_family="L1MB8"
     repeat_region   complement(14096. .14579)
                     /rpt_family="MLT1D"
     repeat_region   complement(14672. .15857)
                     /rpt_family="L1MC1"
     repeat_region   16035. .16060
                     /rpt_family="(TTTTA)n"
     repeat_region   complement(16257. .16866)
                     /rpt_family="MER92B"
     repeat_region   complement(17022. .17865)
                     /rpt_family="L1MC1"
     repeat_region   17943. .17970
                     /rpt_family="AT_rich"
     repeat_region   complement(18142. .18523)
                     /rpt_family="MER44D"
     repeat_region   18524. .18722
                     /rpt_family="LTR28"
     repeat_region   18753. .18800
                     /rpt_family="MER52A"
     repeat_region   18864. .18945
                     /rpt_family="MER52A"
     repeat_region   18921. .19174
                     /rpt_family="MER52A"
     repeat_region   complement(19266. .19418)
                     /rpt_family="MER44D"
     repeat_region   19496. .19533
                     /rpt_family="(TG)n"
     repeat_region   19947. .19976
                     /rpt_family="AT_rich"
     repeat_region   complement(21140. .21266)
                     /rpt_family="MIR"
     repeat_region   22399. .22448
                     /rpt_family="AT_rich"
     repeat_region   complement(22663. .22801)
                     /rpt_family="L2"
     repeat_region   complement(22900. .23196)
                     /rpt_family="MLT1J2"
     repeat_region   complement(23259. .23360)
                     /rpt_family="L2"
     repeat_region   23396. .23696
                     /rpt_family="AluSq"
     repeat_region   23798. .24238
                     /rpt_family="L2"
     repeat_region   complement(24251. .24492)
                     /rpt_family="MIR"
     repeat_region   26206. .26233
                     /rpt_family="AT_rich"
     repeat_region   complement(26553. .26653)
                     /rpt_family="L2"
     repeat_region   complement(26884. .29591)
                     /rpt_family="L1PA8"

  Query Match            65.0%;  Score 23.4;  DB 9;  Length 137154;
  Best Local Similarity  81.8%;  Pred. No. 35;
  Matches   27;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      4 aaagcttccaccatgagccaggacaccaaagtg 36
          ||||||||||| ||| || |||  ||| |||||
Db  50002 AAAGCTTCCACAATGTGCAAGGGGACCCAAGTG 50034


RESULT  14
HSU66464
LOCUS       HSU66464               2732 bp    mRNA    linear   PRI 30-JAN-1998
```

```
DEFINITION  Human hematopoietic progenitor kinase (HPK1) mRNA, complete cds.
ACCESSION   U66464
VERSION     U66464.1  GI:1575562
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2732)
  AUTHORS   Hu,M.C., Qiu,W.R., Wang,X., Meyer,C.F. and Tan,T.H.
  TITLE     Human HPK1, a novel human hematopoietic progenitor kinase that
            activates the JNK/SAPK kinase cascade
  JOURNAL   Genes Dev. 10 (18), 2251-2264 (1996)
  MEDLINE   96421968
REFERENCE   2  (bases 1 to 2732)
  AUTHORS   Hu,M.C.-T., Qiu,W.R., Wang,X., Meyer,C.F. and Tan,T.-H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-AUG-1996) 14-1-D, Amgen, Inc., 1840 DeHavilland
            Drive, Thousand Oaks, CA 91320, USA
FEATURES             Location/Qualifiers
     source          1. .2732
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            1. .2732
                     /gene="HPK1"
     CDS             55. .2556
                     /gene="HPK1"
                     /note="serine/threonine protein kinase"
                     /codon_start=1
                     /product="hematopoietic progenitor kinase"
                     /protein_id="AAB97983.1"
                     /db_xref="GI:1575563"
                     /translation="MDVVDPDIFNRDPRDHYDLLQRLGGGTYGEVFKARDKVSGDLVA
                     LKMVKMEPDDDVSTLQKEILILKTCRHANIVAYHGSYLWLQKLWICMEFCGAGSLQDI
                     YQVTGSLSELQISYVCREVLQGLAYLHSQKKIHRDIKGANILLINDAGEVRLADFGISA
                     QIGATLARRLSFIGTPYWMAPEVAAVALKGGYNELCDIWSLGITAIELAELQPPLFDV
                     HPLRVLFLMTKSGYQPPRLKEKGKWSAAFHNFIKVTLTKSPKKRPSATKMLSHQLVSQ
                     PGLNRGLILDLLDKLKNPGKGPSIGDIEDEEPELPPAIPRRIRSTHRSSSLGIPDADC
                     CRRHMEFRKLRGMETRPPANTARLQPPRDLRSSSPRKQLSESSDDDYDDVDIPTPAED
                     TPPPLPPKPKFRSPSDEGPGSMGDDGQLSPGVLVRCASGPPPNSPRPGPPPSTSSPHL
                     TAHSEPSLWNPPSRELDKPPLLPPKKEKMKRKGCALLVKLFNGCPLRIHSTAAWTHPS
                     TKDQHLLLGAEEGIFILNRNDQEATLEMLFPSRTTWVYSINNVLMSLSGKTPHLYSHS
                     ILGLLERKETRAGNPIAHISPHRLLARKNMVSTKIQDTKGCRACCVAEGASSGGPFLC
                     GALETSVVLLQWYQPMNKFLLVRQVLFPLPTPLSVFALLTGPGSELPAVCIGVSPGRP
                     GKSVLFHTVRFGALSCWLGEMSTEHRGPVQVTQVEEDMVMVLMDGSVKLVTPEGSPVR
                     GLRTPEIPMTEAVEAVAMVGGQLQAFWKHGVQVWALGSDQLLQELRDPTLTFRLLGSP
                     RLECSGTISPHCNLLLPGSSNSPASASRVAGITGL"
BASE COUNT      584 a    848 c    759 g    541 t
ORIGIN
```

```
  Query Match             64.4%;  Score 23.2;  DB 9;  Length 2732;
  Best Local Similarity   89.3%;  Pred. No. 71;
  Matches   25;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      7 gcttccaccatgagccaggacaccaaag 34
          | ||||||||  || ||||||||||||||
Db   1843 GTTTCCACCAAGATCCAGGACACCAAAG 1870
```

```
RESULT  15
AC008649
LOCUS       AC008649              126682 bp    DNA     linear   PRI 03-OCT-2001
DEFINITION  Homo sapiens chromosome 19 clone CTB-186G2, complete sequence.
ACCESSION   AC008649
VERSION     AC008649.7  GI:15887247
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 126682)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 126682)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 126682)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-AUG-2000) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   4  (bases 1 to 126682)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-OCT-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Oct 3, 2001 this sequence version replaced gi:9690315.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.8% of Sequence;
            Estimated Total Number of Errors is 0.2.
            STS Content:
            WI-17325 G24543
            SHGC-8810 G14342.
FEATURES             Location/Qualifiers
     source          1. .126682
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="CTB-186G2"
BASE COUNT    31351 a   33443 c   31369 g   30519 t
ORIGIN
```

```
  Query Match             64.4%;  Score 23.2;  DB 9;  Length 126682;
  Best Local Similarity   89.3%;  Pred. No. 43;
  Matches   25;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      7 gcttccaccatgagccaggacaccaaag 34
          | ||||||||  || ||||||||||||||
Db 113459 GTTTCCACCAAGATCCAGGACACCAAAG 113486
```

```
Search completed: June 20, 2002, 14:35:35
Job time: 19598 sec
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2002, 14:39:36 ; Search time 668.44 Seconds
(without alignments)
92.467 Million cell updates/sec

Title: US-09-927-267-12
Perfect score: 36
Sequence: 1 gcgaaagcttccaccatgagccaggacaccaaagtg 36

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	24	66.7	1835	24	ABA04602	MOL10a coding sequ
	2	24	66.7	2551	24	ABA04603	MOL10b coding sequ
	3	23.6	65.6	1852	22	AAH74768	Nucleotide sequenc
	4	23.6	65.6	1861	24	ABI99366	Mouse ischaemic co
	5	23.2	64.4	3143	23	AAS65625	DNA encoding novel
c	6	23.2	64.4	4266	22	AAS44980	cDNA encoding nove
c	7	23.2	64.4	30393	22	AAK67239	Human immune/haema
	8	22.4	62.2	1797	21	AAA08375	Rat 4F2 heavy chai
	9	22.4	62.2	1940	22	AAF77277	Rat cDNA encoding
	10	22	61.1	1863	21	AAA08374	Human 4F2 heavy ch
	11	22	61.1	1897	22	AAH74770	Nucleotide sequenc
	12	22	61.1	1897	22	AAF77279	Human cDNA encodin
	13	22	61.1	1964	24	AAS94775	Human DNA sequence
c	14	21.8	60.6	493	22	AAK65773	Human immune/haema
	15	21	58.3	44400	22	AAS12438	DNA encoding 1-ami
c	16	20.8	57.8	494	22	AAH04847	Human cDNA clone (
c	17	20.8	57.8	2162	22	AAH17498	Human cDNA sequenc
c	18	20.6	57.2	1982	20	AAZ90535	C86b 1-FFT enzyme
c	19	20.4	56.7	1140	21	AAF13008	Aspergillus oryzae
	20	20.4	56.7	1911	23	AAS67745	DNA encoding novel
	21	20.4	56.7	2029	21	AAA78540	Plant SDF polynucl
c	22	20.4	56.7	3275	22	AAK53220	Human polynucleoti
c	23	20.4	56.7	4325	23	AAS83357	DNA encoding novel
	24	20.2	56.1	272	22	ABA18332	Human nervous syst
	25	20.2	56.1	274	22	ABA12812	Human nervous syst
c	26	20.2	56.1	313	22	ABA08932	Human secreted pro
	27	20.2	56.1	472	22	AAI93186	Human polynucleoti
c	28	20.2	56.1	791	22	AAH04592	Human cDNA clone (
	29	20.2	56.1	954	19	AAX75763	Human microtubule
	30	20.2	56.1	17073	21	AAF20998	Human low adenosin
	31	20.2	56.1	17073	21	AAA34876	Human adenosine re
	32	20.2	56.1	19883	21	AAF20999	Human low adenosin
	33	20.2	56.1	19943	21	AAA34877	Human adenosine re
c	34	20	55.6	4312	22	AAF25736	Tobacco invertase
	35	19.8	55.0	65	24	AAI70936	Influenza virus M2
	36	19.8	55.0	104	22	ABA74073	Human foetal liver
	37	19.8	55.0	104	22	AAK22530	Human brain expres
	38	19.8	55.0	104	22	AAK48697	Human bone marrow
	39	19.8	55.0	104	22	AAI54527	Probe #23213 used
	40	19.8	55.0	585	22	ABA61576	Human foetal liver
	41	19.8	55.0	585	22	AAK09880	Human brain expres
	42	19.8	55.0	585	22	AAK35773	Human bone marrow
	43	19.8	55.0	585	22	AAI41490	Probe #10176 used
c	44	19.8	55.0	4126	22	AAS61020	Human cancer agent
	45	19.8	55.0	4622	24	AAI70939	Vaccine vector pM2

ALIGNMENTS

RESULT 1
ABA04602
ID ABA04602 standard; cDNA; 1835 BP.
XX
AC ABA04602;
XX
DT 21-FEB-2002 (first entry)
XX
DE MOL10a coding sequence.
XX
KW MOL; G-coupled protein-receptor; cardiomyopathy; atherosclerosis;
KW cell signal processing; metabolic disorder; diabetes; cancer;
KW neurodegenerative disorder; immune disorder; cardiac disorder;
KW lung disease; autoimmune disease; developmental disorder; antidiabetic;
KW Cytostatic; Neuroprotective; Antiatherosclerotic; Immunosuppressive;
KW Gene therapy; Vaccine; antiinflammatory; MOL10a; ss.
XX
OS Unidentified.
XX
FH Key Location/Qualifiers
FT CDS 54..1790
FT /*tag= a
FT /product= "MOL 10a"
XX
PN WO200181578-A2.
XX
PD 01-NOV-2001.
XX
PF 26-APR-2001; 2001WO-US13578.
XX
PR 26-APR-2000; 2000US-200158P.
PR 28-APR-2000; 2000US-200613P.

PR 28-APR-2000; 2000US-200780P.
PR 01-MAY-2000; 2000US-201006P.
PR 01-MAY-2000; 2000US-201007P.
PR 01-MAY-2000; 2000US-201236P.
PR 01-MAY-2000; 2000US-201238P.
PR 02-MAY-2000; 2000US-201186P.
PR 03-MAY-2000; 2000US-201474P.
PR 03-MAY-2000; 2000US-201508P.
PR 25-JUL-2000; 2000US-220591P.
PR 15-SEP-2000; 2000US-232678P.
PR 22-JAN-2001; 2001US-263217P.
PR 30-JAN-2001; 2001US-265160P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Vernet CAM, Fernandes ER, Gerlach V, Shimkets RA, Malyankar UM;
PI Boldog FL, Zerhusen BD, Spytek KA, Majumder K, Tchernev VT;
PI Padigaru M, Patturajan M, Burgess CE, Gangolli EA, Smithson G;
PI Rastelli L, Macdougall JR, Taupier RJ, Grosse WM, Szekeres ES;
PI Alsobrook JP;
XX
DR WPI; 2002-049278/06.
DR P-PSDB; AAM47672.
XX
PT Novel G-protein coupled receptor-related polypeptides and
PT polynucleotides for diagnosing, preventing and treating cardiomyopathy,
PT atherosclerosis, disorders related to cell signal processing and for
PT identifying modulators -
XX
PS Claim 8; Page 76; 227pp; English.
XX
CC The present invention relates to novel G-coupled protein-receptor related
CC proteins and coding sequences (MOLX, where X is a number from 1 to 10,
CC ABA04589-ABA04603 and AAM47659-AAM47673). MOLX proteins and coding
CC sequences are useful for treating or preventing a MOLX-associated
CC disorder, such as cardiomyopathy, atherosclerosis, disorders related to
CC cell signal processing and metabolic pathway modulation, diabetes and
CC cancer. Additionally, MOLX proteins and coding sequences are useful for
CC preventing and treating a variety of disorders including metabolic
CC disorders, nutritional oedema, chronic and hereditary pancreatitis,
CC obesity, infectious disease, anorexia, neurodegenerative disorders,
CC Alzheimer's disease, Parkinson's disease, stroke, immune disorders,
CC haematopoietic disorders and various dyslipidaemias, metabolic syndrome X
CC and wasting disorders associated with chronic diseases and cancers,
CC cardiac disorders, hypertension, hypercalcaemia, cirrhosis, angiogenesis
CC and wound healing, trauma, glomerulonephritis, hyper and hypothyroidism,
CC multiple sclerosis, lung diseases including asthma, Crohn's disease,
CC scleroderma, autoimmune diseases, developmental disorders and neural tube
CC defects. The present sequence is the coding sequence for MOL10a.
CC MOL10a protein is a cyclic nucleotide-gated olfactory channel-like
CC protein.
XX
SQ Sequence 1835 BP; 420 A; 521 C; 521 G; 373 T; 0 other;

Query Match 66.7%; Score 24; DB 24; Length 1835;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 accatgagccaggacaccaaagtg 36
||||||||||||||||||||||||
Db 51 accatgagccaggacaccaaagtg 74

RESULT 2
ABA04603
ID ABA04603 standard; cDNA; 2551 BP.
XX
AC ABA04603;
XX
DT 21-FEB-2002 (first entry)
XX
DE MOL10b coding sequence.
XX
KW MOL; G-coupled protein-receptor; cardiomyopathy; atherosclerosis;
KW cell signal processing; metabolic disorder; diabetes; cancer; human;
KW neurodegenerative disorder; immune disorder; cardiac disorder;
KW lung disease; autoimmune disease; developmental disorder; antidiabetic;
KW Cytostatic; Neuroprotective; Antiatherosclerotic; Immunosuppressive;
KW Gene therapy; Vaccine; antiinflammatory; MOL10b; chromosome 11; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 779..2506
FT /*tag= a
FT /product= "MOL 10b"
XX
PN WO200181578-A2.
XX
PD 01-NOV-2001.
XX
PF 26-APR-2001; 2001WO-US13578.
XX
PR 26-APR-2000; 2000US-200158P.
PR 28-APR-2000; 2000US-200613P.
PR 28-APR-2000; 2000US-200780P.
PR 01-MAY-2000; 2000US-201006P.
PR 01-MAY-2000; 2000US-201007P.
PR 01-MAY-2000; 2000US-201236P.
PR 01-MAY-2000; 2000US-201238P.
PR 02-MAY-2000; 2000US-201186P.
PR 03-MAY-2000; 2000US-201474P.
PR 03-MAY-2000; 2000US-201508P.
PR 25-JUL-2000; 2000US-220591P.
PR 15-SEP-2000; 2000US-232678P.
PR 22-JAN-2001; 2001US-263217P.
PR 30-JAN-2001; 2001US-265160P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Vernet CAM, Fernandes ER, Gerlach V, Shimkets RA, Malyankar UM;
PI Boldog FL, Zerhusen BD, Spytek KA, Majumder K, Tchernev VT;
PI Padigaru M, Patturajan M, Burgess CE, Gangolli EA, Smithson G;
PI Rastelli L, Macdougall JR, Taupier RJ, Grosse WM, Szekeres ES;
PI Alsobrook JP;
XX
DR WPI; 2002-049278/06.
DR P-PSDB; AAM47673.
XX
PT Novel G-protein coupled receptor-related polypeptides and
PT polynucleotides for diagnosing, preventing and treating cardiomyopathy,
PT atherosclerosis, disorders related to cell signal processing and for
PT identifying modulators -
XX
PS Claim 8; Page 78; 227pp; English.
XX
CC The present invention relates to novel G-coupled protein-receptor related
CC proteins and coding sequences (MOLX, where X is a number from 1 to 10,
CC ABA04589-ABA04603 and AAM47659-AAM47673). MOLX proteins and coding
CC sequences are useful for treating or preventing a MOLX-associated
CC disorder, such as cardiomyopathy, atherosclerosis, disorders related to
CC cell signal processing and metabolic pathway modulation, diabetes and
CC cancer. Additionally, MOLX proteins and coding sequences are useful for
CC preventing and treating a variety of disorders including metabolic
CC disorders, nutritional oedema, chronic and hereditary pancreatitis,
CC obesity, infectious disease, anorexia, neurodegenerative disorders,
CC Alzheimer's disease, Parkinson's disease, stroke, immune disorders,
CC haematopoietic disorders and various dyslipidaemias, metabolic syndrome X
CC and wasting disorders associated with chronic diseases and cancers,
CC cardiac disorders, hypertension, hypercalcaemia, cirrhosis, angiogenesis
CC and wound healing, trauma, glomerulonephritis, hyper and hypothyroidism,
CC multiple sclerosis, lung diseases including asthma, Crohn's disease,
CC scleroderma, autoimmune diseases, developmental disorders and neural tube
CC defects. The present sequence is the coding sequence for human MOL10b,
CC which maps to chromosome 11. MOL10b protein is a cyclic nucleotide-gated

```
CC   olfactory channel-like protein.
XX
SQ   Sequence 2551 BP; 646 A; 668 C; 707 G; 530 T; 0 other;


  Query Match            66.7%;  Score 24;  DB 24;  Length 2551;
  Best Local Similarity  100.0%;  Pred. No. 2.3;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     13 accatgagccaggacaccaaagtg 36
          ||||||||||||||||||||||||
Db    776 accatgagccaggacaccaaagtg 799


RESULT   3
AAH74768
ID   AAH74768 standard; DNA; 1852 BP.
XX
AC   AAH74768;
XX
DT   29-OCT-2001  (first entry)
XX
DE   Nucleotide sequence of a murine polypeptide.
XX
KW   Sodium independent small neutral amino acid transporter; cell resistance;
KW   oxidizing stress; NMDA type glutamate receptor; cell growth; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             106..1686
FT                   /*tag=  a
XX
PN   WO200158944-A1.
XX
PD   16-AUG-2001.
XX
PF   09-JAN-2001; 2001WO-JP00031.
XX
PR   07-FEB-2000; 2000JP-0028822.
XX
PA   (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
PI   Endou H,  Kanai Y;
XX
DR   WPI; 2001-522461/57.
DR   P-PSDB; AAG63833.
XX
PT   Sodium independent small neutral amino acid transporting proteins -
XX
PS   Disclosure; Page 57-62; 73pp; Japanese.
XX
CC   The specification describes sodium independent small neutral amino
CC   acid transporter polypeptides. The polypeptides and polynucleotides
CC   are used for controlling cell resistance to oxidizing stress, NMDA type
CC   glutamate receptor activity, cell growth and for screening for
CC   substances that have these activities. The present sequence encodes
CC   a polypeptide which is used in the course of the invention.
XX
SQ   Sequence 1852 BP; 417 A; 510 C; 503 G; 422 T; 0 other;


  Query Match            65.6%;  Score 23.6;  DB 22;  Length 1852;
  Best Local Similarity  86.7%;  Pred. No. 3.2;
  Matches   26;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      7 gcttccaccatgagccaggacaccaaagtg 36
          ||   ||||||||||||||||||| |||||
Db     97 gcaggcaccatgagccaggacaccgaagtg 126


RESULT   4
ABI99366
ID   ABI99366 standard; cDNA; 1861 BP.
XX
AC   ABI99366;
XX
DT   07-MAR-2002  (first entry)
XX
DE   Mouse ischaemic condition related cDNA sequence SEQ ID NO:260.
XX
KW   Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;
KW   vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.
XX
OS   Mus musculus.
XX
PN   WO200188188-A2.
XX
PD   22-NOV-2001.
XX
PF   18-MAY-2001; 2001WO-JP04192.
XX
PR   18-MAY-2000; 2000JP-0145977.
XX
PA   (UYNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.
XX
PI   Ishikawa K,  Asai S,  Takahashi Y,  Nagata T,  Ishii Y;
XX
DR   WPI; 2002-034733/04.
DR   P-PSDB; ABB57115.
XX
PT   Examining the ischemic condition (e.g. occlusive ischemia) by measuring
PT   expression levels of particular genes defined in the specification or
PT   by determining the expression profile of a gene group comprising these
PT   genes -
XX
PS   Claim 2; Page 734-737; 2690pp; English.
XX
CC   The present invention describes a method for examining ischaemic
CC   conditions, comprising measuring the expression levels of particular
CC   genes (I) in a test sample or determining the expression profile of a
CC   gene group in the sample comprising genes selected from (I). The method
CC   is useful for examining the ischaemic condition (e.g. compressive
CC   ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring
CC   expression levels of particular genes (ABI99202 to ABI99912, encoding
CC   the protein sequences in ABB57020 to ABB57374) or by determining the
CC   expression profile of a gene group comprising these genes. The
CC   expression levels or expression profiles produced by these genes are
CC   used as an indicator when screening for ischaemic condition-improving
CC   drugs or therapeutics for ischaemic diseases. ABI99913 and ABI99914
CC   represent PCR primers for a mouse ischaemic condition related sequence,
CC   which are used in the exemplification of the present invention.
XX
SQ   Sequence 1861 BP; 404 A; 519 C; 520 G; 418 T; 0 other;


  Query Match            65.6%;  Score 23.6;  DB 24;  Length 1861;
  Best Local Similarity  86.7%;  Pred. No. 3.2;
  Matches   26;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      7 gcttccaccatgagccaggacaccaaagtg 36
          ||   ||||||||||||||||||| |||||
Db    101 gcaggcaccatgagccaggacaccgaagtg 130


RESULT   5
AAS65625
ID   AAS65625 standard; cDNA; 3143 BP.
XX
AC   AAS65625;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #1429.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
```

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG01438.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 1429; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3143 BP; 662 A; 951 C; 857 G; 673 T; 0 other;

Query Match 64.4%; Score 23.2; DB 23; Length 3143;
Best Local Similarity 89.3%; Pred. No. 5.2;
Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 7 gcttccaccatgagccaggacaccaaag 34
| |||||||| || ||||||||||||
Db 1843 gtttccaccaagatccaggacaccaaag 1870

RESULT 6
AAS44980/c
ID AAS44980 standard; cDNA; 4266 BP.
XX
AC AAS44980;
XX
DT 18-DEC-2001 (first entry)
XX
DE cDNA encoding novel human secretory protein, Seq ID No 61.
XX
KW Human; secreted protein; arthritis; Crohn's disease; sepsis; shock;
KW ischaemia-reperfusion injury; haematopoiesis; cancer; neuropathy;
KW transgenic animal; Alzheimer's disease; Parkinson's disease; burn;
KW amyotrophic lateral sclerosis; platelet disorder; thrombocytopenia;
KW ulcer; osteoporosis; bone degenerative disorder; periodontal disease;
KW gut protection; lung; liver fibrosis; immune deficiency; infection;
KW severe combined immunodeficiency; SCID; autoimmune disorder; allergy;
KW multiple sclerosis; rheumatoid arthritis; diabetes mellitus; asthma;
KW fertility; analgesic; pain; antigen; ss.
XX
OS Homo sapiens.
XX
PN WO200166689-A2.
XX
PD 13-SEP-2001.
XX
PF 05-MAR-2001; 2001WO-US04942.
XX
PR 07-MAR-2000; 2000US-0519705.
PR 19-MAY-2000; 2000US-0574454.
PR 17-JUN-2000; 2000US-0596193.
PR 14-JUL-2000; 2000US-0616847.
PR 19-SEP-2000; 2000US-0665363.
PR 20-OCT-2000; 2000US-0693267.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Xu C, Wehrman T, Ren F, Ma Y, Zhou P;
PI Zhao QA, Yang Y, Drmanac RT, Zhang J, Chen R, Xue AJ, Wang J;
XX
DR WPI; 2001-589934/66.
DR P-PSDB; AAU28080.
XX
PT Novel polypeptides and nucleic acids obtained from cDNA libraries
PT prepared from various human tissues, for diagnosis and treatment of
PT cancer, neurological, inflammatory, and autoimmune disorders -
XX
PS Claim 1; SEQ ID No 61; 107pp; English.
XX
CC The invention relates to novel isolated human secreted polypeptides (I)
CC and polynucleotides (II). (I) and (II) are useful for treating
CC inflammatory conditions such as arthritis, nephritis, Crohn's disease,
CC ischaemia-reperfusion injury, shock, sepsis, immune responses, and is
CC involved in increasing haematopoiesis, stem cell survival, bone growth
CC and remodeling. (I), (II) and modulators of (II) are useful for
CC prophylaxis or treatment of one or more cancers. (II) is also useful for
CC creating transgenic animals useful for studying the in vivo activities of
CC the polypeptide as well as for studying modulators of the polypeptides.
CC (I) induces the proliferation of neural cells and regeneration of nerve
CC and brain tissue and is useful for the treatment of central and
CC peripheral nervous system diseases and neuropathies, such as Alzheimer's,
CC Parkinson's disease, Huntington's disease, and amyotrophic lateral
CC sclerosis. In addition, (I) is involved in chemotactic or chemokinetic
CC activity, regulation of haematopoiesis and is useful for treating myeloid
CC or lymphoid cell disorders, platelet disorders such as thrombocytopenia
CC and for regeneration of bone, cartilage, tendon, ligament and/or nerve
CC tissue growth, and in tissue repair, healing of burns, incisions,
CC ulcers, for treating osteoporosis, osteoarthritis, bone degenerative
CC disorders, or periodontal disease. Furthermore, (I) is also useful for
CC gut protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues, various immune deficiencies and
CC disorders including severe combined immunodeficiency (SCID), bacterial or
CC fungal infections, autoimmune disorders e.g. multiple sclerosis,
CC rheumatoid arthritis, diabetes mellitus, myasthenia gravis, allergic
CC reactions and conditions, such as asthma or other respiratory problems.
CC In addition, (I) affects biorhythms or circadian cycles of rhythms,
CC fertility, metabolism, catabolism, anabolism, storage or elimination of
CC dietary fat, lipid, protein, carbohydrate, vitamins, minerals, provides
CC analgesic effects or other pain reducing effects, immunoglobulin like
CC activity and can act as an antigen in a vaccine composition to raise an
CC immune response. AAS44920-AAS45295 represent novel human secreted protein
CC coding sequences of the invention.
XX
SQ Sequence 4266 BP; 1107 A; 1077 C; 1172 G; 910 T; 0 other;

```
  Query Match             64.4%;  Score 23.2;  DB 22;  Length 4266;
  Best Local Similarity   89.3%;  Pred. No. 5.6;
  Matches   25;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      7 gcttccaccatgagccaggacaccaaag 34
          | |||||||| || ||||||||||||||
Db   2478 GTTTCCACCAAGATCCAGGACACCAAAG 2451


RESULT   7
AAK67239/c
ID   AAK67239 standard; DNA; 30393 BP.
XX
AC   AAK67239;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:22051.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
```

```
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Disclosure; SEQ ID NO 22051; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 30393 BP; 8271 A; 7315 C; 8076 G; 6731 T; 0 other;


  Query Match             64.4%;  Score 23.2;  DB 22;  Length 30393;
  Best Local Similarity   89.3%;  Pred. No. 8;
  Matches   25;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       7 gcttccaccatgagccaggacaccaaag 34
            | |||||||| || ||||||||||||||
Db    9748 GTTTCCACCAAGATCCAGGACACCAAAG 9721


RESULT   8
AAA08375
ID   AAA08375 standard; cDNA; 1797 BP.
XX
AC   AAA08375;
XX
DT   12-JUL-2000  (first entry)
XX
DE   Rat 4F2 heavy chain nucleotide sequence SEQ ID NO:7.
XX
KW   L-type amino acid transporter 1; LAT1; 4F2hc; 4F2 heavy chain;
KW   neutral amino acid transporter; tumour; cancer; proliferation;
KW   cell membrane surface 4F2 molecule; anticancer; ss.
XX
OS   Rattus sp.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1..19
FT                   /*tag= a
FT   CDS             20..1603
FT                   /*tag= b
FT                   /product= "4F2hc"
FT   3'UTR           1604..1797
FT                   /*tag= c
XX
PN   WO200014228-A1.
XX
PD   16-MAR-2000.
XX
PF   03-SEP-1999;   99WO-JP04789.
XX
PR   03-SEP-1998;   98JP-0249993.
PR   02-SEP-1999;   99JP-0248546.
XX
PA   (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
PI   Endou H,  Kanai Y;
XX
DR   WPI; 2000-256979/22.
DR   P-PSDB; AAY82490.
XX
PT   Neutral amino acid transporter protein which conjugates with cell
PT   membrane protein 4F2 and operates independently of sodium ions is
PT   useful for screening potential cancer proliferation inhibitors -
XX
PS   Example 1; Page 157-162; 189pp; Japanese.
XX
CC   The present invention describes a cell surface protein which is an
CC   L-type amino acid transporter-1 (LAT1), which mediates the transport
CC   of neutral amino acids, leucine, isoleucine, phenylalanine, methionine,
CC   tyrosine, tryptophan, valine and histidine, into the cell independently
CC   of sodium ions. The LAT1 protein conjugates with the cell membrane
CC   surface molecule 4F2. Antisense DNA or RNA, antibodies, antibody
CC   fragments, and peptide and non-peptide antagonists to LAT1 are useful
CC   as anticancer agents. The present sequence encodes rat 4F2 heavy chain
CC   (4F2hc), which is specifically claimed in the present invention.
XX
SQ   Sequence 1797 BP; 429 A; 466 C; 478 G; 424 T; 0 other;


  Query Match             62.2%;  Score 22.4;  DB 21;  Length 1797;
  Best Local Similarity   95.8%;  Pred. No. 10;
  Matches   23;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      13 accatgagccaggacaccaaagtg 36
           |||||||||||||||||| |||||
Db      17 accatgagccaggacaccgaagtg 40


RESULT   9
AAF77277
ID   AAF77277 standard; cDNA; 1940 BP.
XX
AC   AAF77277;
XX
DT   24-MAY-2001  (first entry)
XX
DE   Rat cDNA encoding amino acid transport activator 4F2 heavy chain 4F2hc.
XX
KW   Neutral amino acid transporter; amino acid transport activating factor;
KW   4F2hc; 4F2 heavy chain; drug permeation; poison; rat; LAT2; ss;
KW   L-type amino acid transporter 2.
```

```
XX
OS   Rattus sp.
XX
PN   JP2000342270-A.
XX
PD   12-DEC-2000.
XX
PF   04-JUN-1999;   99JP-0158571.
XX
PR   04-JUN-1999;   99JP-0158571.
XX
PA   (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
XX
DR   WPI; 2001-260288/27.
DR   P-PSDB; AAB72394.
XX
PT   A neutral amino acid transporter having broad substrate selectivity
PT   useful for in vitro analysis of permeation of drugs and poisons through
PT   cell membranes, blood and tissues -
XX
PS   Example 1; Page 18-21; 37pp; Japanese.
XX
CC   This invention relates to a neutral amino acid transporter having broad
CC   substrate selectivity to neutral amino acids, and which is inactivated by
CC   an amino acid transport activating factor 4F2hc (4F2 heavy chain). The
CC   neutral amino acid transporter and its gene are useful in in vitro
CC   analysis of permeation of drugs and poisons through cell membrane, blood
CC   and tissues. The present sequence represents cDNA encoding rat amino acid
CC   transport activator 4F2 heavy chain (4F2hc).
XX
SQ   Sequence 1940 BP; 431 A; 534 C; 517 G; 458 T; 0 other;


  Query Match             62.2%;  Score 22.4;  DB 22;  Length 1940;
  Best Local Similarity   95.8%;  Pred. No. 10;
  Matches   23;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      13 accatgagccaggacaccaaagtg 36
           |||||||||||||||||| |||||
Db     194 accatgagccaggacaccgaagtg 217


RESULT  10
AAA08374
ID   AAA08374 standard; cDNA; 1863 BP.
XX
AC   AAA08374;
XX
DT   12-JUL-2000  (first entry)
XX
DE   Human 4F2 heavy chain nucleotide sequence SEQ ID NO:5.
XX
KW   L-type amino acid transporter 1; LAT1; 4F2hc; 4F2 heavy chain;
KW   neutral amino acid transporter; tumour; cancer; proliferation;
KW   cell membrane surface 4F2 molecule; anticancer; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1..109
FT                   /*tag=  a
FT   CDS             110..1699
FT                   /*tag=  b
FT                   /product= "4F2hc"
FT   3'UTR           1700..1863
FT                   /*tag=  c
XX
PN   WO200014228-A1.
XX
PD   16-MAR-2000.
XX
PF   03-SEP-1999;   99WO-JP04789.
XX
PR   03-SEP-1998;   98JP-0249993.
PR   02-SEP-1999;   99JP-0248546.
XX
PA   (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
PI   Endou H,  Kanai Y;
XX
DR   WPI; 2000-256979/22.
DR   P-PSDB; AAY82490.
XX
PT   Neutral amino acid transporter protein which conjugates with cell
PT   membrane protein 4F2 and operates independently of sodium ions is
PT   useful for screening potential cancer proliferation inhibitors -
XX
PS   Claim 16; Page 148-153; 189pp; Japanese.
XX
CC   The present invention describes a cell surface protein which is an
CC   L-type amino acid transporter-1 (LAT1), which mediates the transport
CC   of neutral amino acids, leucine, isoleucine, phenylalanine, methionine,
CC   tyrosine, tryptophan, valine and histidine, into the cell independently
CC   of sodium ions. The LAT1 protein conjugates with the cell membrane
CC   surface molecule 4F2. Antisense DNA or RNA, antibodies, antibody
CC   fragments, and peptide and non-peptide antagonists to LAT1 are useful
CC   as anticancer agents. The present sequence encodes human 4F2 heavy chain
CC   (4F2hc), which is specifically claimed in the present invention.
XX
SQ   Sequence 1863 BP; 370 A; 529 C; 551 G; 413 T; 0 other;


  Query Match             61.1%;  Score 22;  DB 21;  Length 1863;
  Best Local Similarity   83.3%;  Pred. No. 15;
  Matches   25;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       7 gcttccaccatgagccaggacaccaaagtg 36
           ||   |||||||||||||||||| | |||
Db     101 gcaggcaccatgagccaggacaccgaggtg 130


RESULT  11
AAH74770
ID   AAH74770 standard; DNA; 1897 BP.
XX
AC   AAH74770;
XX
DT   29-OCT-2001  (first entry)
XX
DE   Nucleotide sequence of a human polypeptide.
XX
KW   Sodium independent small neutral amino acid transporter; cell resistance;
KW   oxidizing stress; NMDA type glutamate receptor; cell growth; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             112..1701
FT                   /*tag=  a
XX
PN   WO200158944-A1.
XX
PD   16-AUG-2001.
XX
PF   09-JAN-2001; 2001WO-JP00031.
XX
PR   07-FEB-2000; 2000JP-0028822.
XX
PA   (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
PI   Endou H,  Kanai Y;
XX
DR   WPI; 2001-522461/57.
DR   P-PSDB; AAG63805.
XX
PT   Sodium independent small neutral amino acid transporting proteins -
```

```
XX
PS   Disclosure; Page 66-71; 73pp; Japanese.
XX
CC   The specification describes sodium independent small neutral amino
CC   acid transporter polypeptides. The polypeptides and polynucleotides
CC   are used for controlling cell resistance to oxidizing stress, NMDA type
CC   glutamate receptor activity, cell growth and for screening for
CC   substances that have these activities. The present sequence encodes
CC   a polypeptide which is used in the course of the invention.
XX
SQ   Sequence 1897 BP; 390 A; 534 C; 554 G; 419 T; 0 other;


  Query Match             61.1%;  Score 22;  DB 22;  Length 1897;
  Best Local Similarity   83.3%;  Pred. No. 15;
  Matches   25;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       7 gcttccaccatgagccaggacaccaaagtg 36
           ||   |||||||||||||||||| | |||
Db     103 gcaggcaccatgagccaggacaccgaggtg 132


RESULT  12
AAF77279
ID   AAF77279 standard; cDNA; 1897 BP.
XX
AC   AAF77279;
XX
DT   24-MAY-2001  (first entry)
XX
DE   Human cDNA encoding amino acid transport activator 4F2 heavy chain 4F2hc.
XX
KW   Neutral amino acid transporter; amino acid transport activating factor;
KW   4F2hc; 4F2 heavy chain; drug permeation; poison; human; LAT2; ss;
KW   L-type amino acid transporter 2.
XX
OS   Homo sapiens.
XX
PN   JP2000342270-A.
XX
PD   12-DEC-2000.
XX
PF   04-JUN-1999;   99JP-0158571.
XX
PR   04-JUN-1999;   99JP-0158571.
XX
PA   (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
XX
DR   WPI; 2001-260288/27.
DR   P-PSDB; AAB72396.
XX
PT   A neutral amino acid transporter having broad substrate selectivity
PT   useful for in vitro analysis of permeation of drugs and poisons through
PT   cell membranes, blood and tissues -
XX
PS   Example 3; Page 28-31; 37pp; Japanese.
XX
CC   This invention relates to a neutral amino acid transporter having broad
CC   substrate selectivity to neutral amino acids, and which is inactivated by
CC   an amino acid transport activating factor 4F2hc (4F2 heavy chain). The
CC   neutral amino acid transporter and its gene are useful in in vitro
CC   analysis of permeation of drugs and poisons through cell membrane, blood
CC   and tissues. The present sequence represents cDNA encoding human amino
CC   acid transport activator 4F2 heavy chain (4F2hc).
XX
SQ   Sequence 1897 BP; 390 A; 534 C; 554 G; 419 T; 0 other;


  Query Match             61.1%;  Score 22;  DB 22;  Length 1897;
  Best Local Similarity   83.3%;  Pred. No. 15;
  Matches   25;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       7 gcttccaccatgagccaggacaccaaagtg 36
           ||   |||||||||||||||||| | |||
Db     103 gcaggcaccatgagccaggacaccgaggtg 132


RESULT  13
AAS94775
ID   AAS94775 standard; DNA; 1964 BP.
XX
AC   AAS94775;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Human DNA sequence #30 expressed during foam cell differentiation.
XX
KW   Human; foam cell differentiation; atherosclerosis; cerebral stroke;
KW   cardiovascular disorder; coronary artery disease; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200177389-A2.
XX
PD   18-OCT-2001.
XX
PF   04-APR-2001; 2001WO-US11128.
XX
PR   05-APR-2000; 2000US-195106P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Shiffman D,  Somogyi R,  Lawn R,  Seilhamer JJ,  Porter GJ,  Mikita T;
PI   Tai J;
XX
DR   WPI; 2002-010925/01.
XX
PT   Composition useful for diagnosis of conditions, disorders or diseases
PT   associated with atherosclerosis, comprises several polynucleotides that
PT   are differentially expressed in foam cell development -
XX
PS   Claim 1; Page 86; 315pp; English.
XX
CC   The present invention relates to the isolation of human polynucleotide
CC   sequences that are differentially expressed during foam cell
CC   differentiation. The polynucleotide sequences of the invention or a
CC   composition comprising these polynucleotides are useful as a high
CC   throughput method for detecting altered expression of one or more
CC   polynucleotides in a sample. The polynucleotides can be used in the
CC   diagnosis of disorders associated with foam cell development such as
CC   atherosclerosis, cerebral stroke, and cardiovascular disorders such as
CC   coronary artery disease. The polynucleotide sequences can also be used
CC   as PCR primers and probes. The polynucleotides of the invention are also
CC   useful in gene therapy. AAS94746-AAS95021 represent the human
CC   polynucleotide sequences of the invention which are differentially
CC   expressed during foam cell differentiation.
XX
SQ   Sequence 1964 BP; 383 A; 562 C; 586 G; 433 T; 0 other;


  Query Match             61.1%;  Score 22;  DB 24;  Length 1964;
  Best Local Similarity   83.3%;  Pred. No. 15;
  Matches   25;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       7 gcttccaccatgagccaggacaccaaagtg 36
           ||   |||||||||||||||||| | |||
Db     185 gcaggcaccatgagccaggacaccgaggtg 214


RESULT  14
AAK65773/c
ID   AAK65773 standard; DNA; 493 BP.
XX
AC   AAK65773;
XX
DT   06-NOV-2001  (first entry)
```

XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:20585.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.

PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 20585; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 493 BP; 98 A; 88 C; 90 G; 217 T; 0 other;

Query Match 60.6%; Score 21.8; DB 22; Length 493;
Best Local Similarity 78.8%; Pred. No. 14;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 3 gaaagcttccaccatgagccaggacaccaaagt 35
| ||||||| | |||||| | ||||||||| |
Db 266 GGAAGCTTCTAACATGAGAGACGACACCAAAAT 234

RESULT 15
AAS12438
ID AAS12438 standard; DNA; 44400 BP.
XX
AC AAS12438;
XX
DT 04-DEC-2001 (first entry)
XX
DE DNA encoding 1-aminocyclopropane carboxylate synthase #9.
XX
KW 1-aminocyclopropane carboxylate synthase; ACPC synthase; brain trauma;
KW excitatory neurotransmission; neurodegeneration; stroke; nerve damage;
KW neurodegenerative disease; Alzheimer's disease; depression; epilepsy;
KW alcohol abuse; cognitive function; memory; learning impairment;
KW human; ds.
XX
OS Homo sapiens.
XX
PN WO200168879-A2.
XX
PD 20-SEP-2001.
XX
PF 14-MAR-2001; 2001WO-EP02857.
XX
PR 14-MAR-2000; 2000US-0189086.
PR 05-APR-2000; 2000US-0194702.
XX
PA (FARB) BAYER AG.
XX
PI Ramakrishnan S;
XX
DR WPI; 2001-550286/61.
XX
PT Isolated polynucleotide encoding a human
PT 1-aminocyclopropane-carboxylate (ACPC) synthase, useful for treating
PT brain trauma and neurodegenerative disease (e.g. Alzheimer's disease,
PT depression, epilepsy) -
XX
PS Claim 1; Page 196-212; 242pp; English.
XX
CC The invention relates to reagents and methods for regulating excitatory
CC neurotransmission, and to prevent neurodegeneration. The method involves
CC the use of an expression vector or a reagent that modulates the activity
CC of a 1-aminocyclopropane-carboxylate (ACPC) synthase polypeptide. The
CC reagent is useful for modulating the activity of an ACPC synthase in a
CC disease such as stroke, a nerve damage or a neurodegenerative disease.
CC The ACPC synthase polypeptide, polynucleotides and modulators are also
CC useful for treating brain trauma and neurodegenerative disease (e.g.
CC Alzheimer's disease, depression, epilepsy). The ACPC synthase modulators
CC are also useful for treating alcohol abuse and improve cognitive function
CC and memory of patients with learning impairment. The present sequence
CC represents the coding sequence of human 1-aminocyclopropane-carboxylate
CC (ACPC) synthase #9, used in the method of the invention.
XX
SQ Sequence 44400 BP; 11216 A; 9701 C; 10474 G; 13009 T; 0 other;

Query Match 58.3%; Score 21; DB 22; Length 44400;
Best Local Similarity 82.8%; Pred. No. 71;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 4 aaagcttccaccatgagccaggacaccaa 32
|||| || |||||||||||| | ||||
Db 12305 aaagatttcaccatgagccagaaacccaa 12333

Search completed: June 20, 2002, 14:39:39
Job time: 13577 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2002, 14:28:41 ; Search time 155.16 Seconds
(without alignments)
56.992 Million cell updates/sec

Title: US-09-927-267-12
Perfect score: 36
Sequence: 1 gcgaaagcttccaccatgagccaggacaccaaagtg 36

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	20	55.6	840	4	US-08-998-416-507	Sequence 507, App
c	2	19.6	54.4	561	3	US-08-612-973-23	Sequence 23, Appl
c	3	19.6	54.4	561	4	US-08-927-597-23	Sequence 23, Appl
c	4	19.6	54.4	606	3	US-08-612-973-25	Sequence 25, Appl
c	5	19.6	54.4	606	4	US-08-927-597-25	Sequence 25, Appl
c	6	19.6	54.4	636	3	US-08-612-973-27	Sequence 27, Appl
c	7	19.6	54.4	636	4	US-08-927-597-27	Sequence 27, Appl
c	8	19.6	54.4	723	3	US-08-612-973-21	Sequence 21, Appl
c	9	19.6	54.4	723	4	US-08-927-597-21	Sequence 21, Appl
c	10	19.6	54.4	795	3	US-08-612-973-5	Sequence 5, Appli
c	11	19.6	54.4	795	4	US-08-927-597-5	Sequence 5, Appli
c	12	19.6	54.4	2082	3	US-08-612-973-47	Sequence 47, Appl
c	13	19.6	54.4	2082	4	US-08-927-597-47	Sequence 47, Appl
c	14	19.6	54.4	2433	3	US-08-612-973-49	Sequence 49, Appl
c	15	19.6	54.4	2433	4	US-08-927-597-49	Sequence 49, Appl
	16	19.2	53.3	34	2	US-08-765-783A-30	Sequence 30, Appl
	17	19.2	53.3	34	3	US-08-921-100-30	Sequence 30, Appl
	18	19.2	53.3	34	3	US-08-880-142-30	Sequence 30, Appl
	19	19.2	53.3	34	3	US-08-902-201-30	Sequence 30, Appl
	20	19.2	53.3	34	4	US-09-416-557-30	Sequence 30, Appl
c	21	18.8	52.2	3982	3	US-08-947-823-4	Sequence 4, Appli
c	22	18.8	52.2	4149	2	US-08-737-715-1	Sequence 1, Appli
c	23	18.8	52.2	51952	3	US-08-947-823-1	Sequence 1, Appli
	24	18.6	51.7	38	1	US-08-424-424B-5	Sequence 5, Appli
	25	18.6	51.7	55	4	US-08-483-511-14	Sequence 14, Appl
	26	18.6	51.7	55	5	PCT-US93-01009-14	Sequence 14, Appl
c	27	18.6	51.7	284	3	US-08-621-018B-20	Sequence 20, Appl
c	28	18.6	51.7	80246	4	US-09-078-294-4	Sequence 4, Appli
c	29	18.6	51.7	80595	4	US-09-078-294-3	Sequence 3, Appli
c	30	18.6	51.7	84495	4	US-09-797-906-3	Sequence 3, Appli
c	31	18.4	51.1	1214	1	US-08-164-614A-3	Sequence 3, Appli
c	32	18.4	51.1	1214	2	US-08-456-489B-3	Sequence 3, Appli
c	33	18.4	51.1	1214	5	PCT-US93-01720-3	Sequence 3, Appli
c	34	18.4	51.1	1905	1	US-08-164-614A-2	Sequence 2, Appli
c	35	18.4	51.1	1905	2	US-08-456-489B-2	Sequence 2, Appli
c	36	18.4	51.1	1905	5	PCT-US93-01720-2	Sequence 2, Appli
c	37	18.4	51.1	2281	1	US-08-164-614A-1	Sequence 1, Appli
c	38	18.4	51.1	2281	2	US-08-456-489B-1	Sequence 1, Appli
c	39	18.4	51.1	2281	5	PCT-US93-01720-1	Sequence 1, Appli
	40	18.4	51.1	3031	1	US-08-785-241-2	Sequence 2, Appli
	41	18.2	50.6	1448	1	US-08-298-426-1	Sequence 1, Appli
c	42	18.2	50.6	1843	1	US-07-918-023-1	Sequence 1, Appli
c	43	18	50.0	82	2	US-08-476-176B-24	Sequence 24, Appl
c	44	18	50.0	82	3	US-08-127-721A-24	Sequence 24, Appl
c	45	18	50.0	82	3	US-08-485-246A-24	Sequence 24, Appl

ALIGNMENTS

```
RESULT 1
US-08-998-416-507/c
; Sequence 507, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  507:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  840 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1357RP
```

US-08-998-416-507

Query Match 55.6%; Score 20; DB 4; Length 840;
Best Local Similarity 82.1%; Pred. No. 20;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy       6 agcttccaccatgagccaggacaccaaa 33
           || | | |||||||||| |||||||| |
Db     171 AGGTACTACCATGAGCCTGGACACCATA 144
```

```
RESULT   2
US-08-612-973-23/c
; Sequence 23, Application US/08612973
; Patent No. 6150134
;  GENERAL INFORMATION:
;    APPLICANT:  MAERTENS, GEERT
;    APPLICANT:  BOSMAN, FONS
;    APPLICANT:  DE MARTYNOFF, GUY
;    APPLICANT:  BUYSE, MARIE-ANGE
;    TITLE OF INVENTION:  PURIFIED HEPATITIS C VIRUS ENVELOPE
;    TITLE OF INVENTION:  PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
;    NUMBER OF SEQUENCES:  111
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 NORTH GLEBE ROAD
;      CITY:  ARLINGTON
;      STATE:  VIRGINIA
;      COUNTRY:  U.S.A.
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/612,973
;      FILING DATE:  11-MAR-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BYRNE, THOMAS E.
;      REGISTRATION NUMBER:  32,205
;      REFERENCE/DOCKET NUMBER:  1487-10
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 816-4000
;      TELEFAX:  (703) 816-4100
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  561 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..558
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  1..555
US-08-612-973-23
```

Query Match 54.4%; Score 19.6; DB 3; Length 561;
Best Local Similarity 73.5%; Pred. No. 27;
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```
Qy       1 gcgaaagcttccaccatgagccaggacaccaaag 34
           ||| |||||   ||  | || ||||||| |||||
Db     221 GCGGAAGCTGGAACGGTCAGACAGGACAGCAAAG 188
```

```
RESULT   3
US-08-927-597-23/c
; Sequence 23, Application US/08927597
; Patent No. 6245503
;  GENERAL INFORMATION:
;    APPLICANT:  MAERTENS, GEERT
;    APPLICANT:  BOSMAN, FONS
;    APPLICANT:  DE MARTYNOFF, GUY
;    APPLICANT:  BUYSE, MARIE-ANGE
;    TITLE OF INVENTION:  PURIFIED HEPATITIS C VIRUS ENVELOPE
;    TITLE OF INVENTION:  PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
;    NUMBER OF SEQUENCES:  111
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 NORTH GLEBE ROAD
;      CITY:  ARLINGTON
;      STATE:  VIRGINIA
;      COUNTRY:  U.S.A.
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/927,597
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/612,973
;      FILING DATE:  11-MAR-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BYRNE, THOMAS E.
;      REGISTRATION NUMBER:  32,205
;      REFERENCE/DOCKET NUMBER:  1487-10
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 816-4000
;      TELEFAX:  (703) 816-4100
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  561 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..558
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  1..555
US-08-927-597-23
```

Query Match 54.4%; Score 19.6; DB 4; Length 561;
Best Local Similarity 73.5%; Pred. No. 27;
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```
Qy       1 gcgaaagcttccaccatgagccaggacaccaaag 34
           ||| |||||   ||  | || ||||||| |||||
Db     221 GCGGAAGCTGGAACGGTCAGACAGGACAGCAAAG 188
```

```
RESULT   4
US-08-612-973-25/c
; Sequence 25, Application US/08612973
; Patent No. 6150134
;  GENERAL INFORMATION:
;    APPLICANT:  MAERTENS, GEERT
;    APPLICANT:  BOSMAN, FONS
```

```
;    APPLICANT:  DE MARTYNOFF, GUY
;    APPLICANT:  BUYSE, MARIE-ANGE
;    TITLE OF INVENTION:  PURIFIED HEPATITIS C VIRUS ENVELOPE
;    TITLE OF INVENTION:  PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
;    NUMBER OF SEQUENCES:  111
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 NORTH GLEBE ROAD
;      CITY:  ARLINGTON
;      STATE:  VIRGINIA
;      COUNTRY:  U.S.A.
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/612,973
;      FILING DATE:  11-MAR-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BYRNE, THOMAS E.
;      REGISTRATION NUMBER:  32,205
;      REFERENCE/DOCKET NUMBER:  1487-10
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 816-4000
;      TELEFAX:  (703) 816-4100
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  606 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..603
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  1..600
US-08-612-973-25


  Query Match             54.4%;  Score 19.6;  DB 3;  Length 606;
  Best Local Similarity   73.5%;  Pred. No. 27;
  Matches   25;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       1 gcgaaagcttccaccatgagccaggacaccaaag 34
           ||| |||||   ||  | || ||||||| |||||
Db     221 GCGGAAGCTGGAACGGTCAGACAGGACAGCAAAG 188


RESULT   5
US-08-927-597-25/c
; Sequence 25, Application US/08927597
; Patent No. 6245503
;  GENERAL INFORMATION:
;    APPLICANT:  MAERTENS, GEERT
;    APPLICANT:  BOSMAN, FONS
;    APPLICANT:  DE MARTYNOFF, GUY
;    APPLICANT:  BUYSE, MARIE-ANGE
;    TITLE OF INVENTION:  PURIFIED HEPATITIS C VIRUS ENVELOPE
;    TITLE OF INVENTION:  PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
;    NUMBER OF SEQUENCES:  111
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 NORTH GLEBE ROAD
;      CITY:  ARLINGTON
;      STATE:  VIRGINIA
;      COUNTRY:  U.S.A.
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/927,597
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/612,973
;      FILING DATE:  11-MAR-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BYRNE, THOMAS E.
;      REGISTRATION NUMBER:  32,205
;      REFERENCE/DOCKET NUMBER:  1487-10
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 816-4000
;      TELEFAX:  (703) 816-4100
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  606 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..603
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  1..600
US-08-927-597-25


  Query Match             54.4%;  Score 19.6;  DB 4;  Length 606;
  Best Local Similarity   73.5%;  Pred. No. 27;
  Matches   25;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       1 gcgaaagcttccaccatgagccaggacaccaaag 34
           ||| |||||   ||  | || ||||||| |||||
Db     221 GCGGAAGCTGGAACGGTCAGACAGGACAGCAAAG 188


RESULT   6
US-08-612-973-27/c
; Sequence 27, Application US/08612973
; Patent No. 6150134
;  GENERAL INFORMATION:
;    APPLICANT:  MAERTENS, GEERT
;    APPLICANT:  BOSMAN, FONS
;    APPLICANT:  DE MARTYNOFF, GUY
;    APPLICANT:  BUYSE, MARIE-ANGE
;    TITLE OF INVENTION:  PURIFIED HEPATITIS C VIRUS ENVELOPE
;    TITLE OF INVENTION:  PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
;    NUMBER OF SEQUENCES:  111
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 NORTH GLEBE ROAD
;      CITY:  ARLINGTON
;      STATE:  VIRGINIA
;      COUNTRY:  U.S.A.
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/612,973
```

```
;       FILING DATE:  11-MAR-1996
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BYRNE, THOMAS E.
;       REGISTRATION NUMBER:  32,205
;       REFERENCE/DOCKET NUMBER:  1487-10
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703) 816-4000
;       TELEFAX:  (703) 816-4100
;   INFORMATION FOR SEQ ID NO:  27:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  636 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..633
;     FEATURE:
;       NAME/KEY:  mat_peptide
;       LOCATION:  1..630
US-08-612-973-27


  Query Match             54.4%;  Score 19.6;  DB 3;  Length 636;
  Best Local Similarity   73.5%;  Pred. No. 27;
  Matches   25;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       1 gcgaaagcttccaccatgagccaggacaccaaag 34
           ||| |||||   ||  | || ||||||| |||||
Db     221 GCGGAAGCTGGAACGGTCAGACAGGACAGCAAAG 188


RESULT   7
US-08-927-597-27/c
; Sequence 27, Application US/08927597
; Patent No. 6245503
;  GENERAL INFORMATION:
;    APPLICANT:  MAERTENS, GEERT
;    APPLICANT:  BOSMAN, FONS
;    APPLICANT:  DE MARTYNOFF, GUY
;    APPLICANT:  BUYSE, MARIE-ANGE
;    TITLE OF INVENTION:  PURIFIED HEPATITIS C VIRUS ENVELOPE
;    TITLE OF INVENTION:  PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
;    NUMBER OF SEQUENCES:  111
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 NORTH GLEBE ROAD
;      CITY:  ARLINGTON
;      STATE:  VIRGINIA
;      COUNTRY:  U.S.A.
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/927,597
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/612,973
;      FILING DATE:  11-MAR-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BYRNE, THOMAS E.
;      REGISTRATION NUMBER:  32,205
;      REFERENCE/DOCKET NUMBER:  1487-10
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 816-4000
;      TELEFAX:  (703) 816-4100
;  INFORMATION FOR SEQ ID NO:  27:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  636 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..633
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  1..630
US-08-927-597-27


  Query Match             54.4%;  Score 19.6;  DB 4;  Length 636;
  Best Local Similarity   73.5%;  Pred. No. 27;
  Matches   25;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       1 gcgaaagcttccaccatgagccaggacaccaaag 34
           ||| |||||   ||  | || ||||||| |||||
Db     221 GCGGAAGCTGGAACGGTCAGACAGGACAGCAAAG 188


RESULT   8
US-08-612-973-21/c
; Sequence 21, Application US/08612973
; Patent No. 6150134
;  GENERAL INFORMATION:
;    APPLICANT:  MAERTENS, GEERT
;    APPLICANT:  BOSMAN, FONS
;    APPLICANT:  DE MARTYNOFF, GUY
;    APPLICANT:  BUYSE, MARIE-ANGE
;    TITLE OF INVENTION:  PURIFIED HEPATITIS C VIRUS ENVELOPE
;    TITLE OF INVENTION:  PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
;    NUMBER OF SEQUENCES:  111
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 NORTH GLEBE ROAD
;      CITY:  ARLINGTON
;      STATE:  VIRGINIA
;      COUNTRY:  U.S.A.
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/612,973
;      FILING DATE:  11-MAR-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BYRNE, THOMAS E.
;      REGISTRATION NUMBER:  32,205
;      REFERENCE/DOCKET NUMBER:  1487-10
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 816-4000
;      TELEFAX:  (703) 816-4100
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  723 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
```

```
;      NAME/KEY:  CDS
;      LOCATION:  1..720
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  1..717
US-08-612-973-21


  Query Match             54.4%;  Score 19.6;  DB 3;  Length 723;
  Best Local Similarity   73.5%;  Pred. No. 28;
  Matches   25;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      1 gcgaaagcttccaccatgagccaggacaccaaag 34
          ||| |||||   ||  | || ||||||| |||||
Db    221 GCGGAAGCTGGAACGGTCAGACAGGACAGCAAAG 188


RESULT   9
US-08-927-597-21/c
; Sequence 21, Application US/08927597
; Patent No. 6245503
;  GENERAL INFORMATION:
;    APPLICANT:  MAERTENS, GEERT
;    APPLICANT:  BOSMAN, FONS
;    APPLICANT:  DE MARTYNOFF, GUY
;    APPLICANT:  BUYSE, MARIE-ANGE
;    TITLE OF INVENTION:  PURIFIED HEPATITIS C VIRUS ENVELOPE
;    TITLE OF INVENTION:  PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
;    NUMBER OF SEQUENCES:  111
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 NORTH GLEBE ROAD
;      CITY:  ARLINGTON
;      STATE:  VIRGINIA
;      COUNTRY:  U.S.A.
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/927,597
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/612,973
;      FILING DATE:  11-MAR-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BYRNE, THOMAS E.
;      REGISTRATION NUMBER:  32,205
;      REFERENCE/DOCKET NUMBER:  1487-10
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 816-4000
;      TELEFAX:  (703) 816-4100
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  723 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..720
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  1..717
US-08-927-597-21


  Query Match             54.4%;  Score 19.6;  DB 4;  Length 723;
  Best Local Similarity   73.5%;  Pred. No. 28;
  Matches   25;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      1 gcgaaagcttccaccatgagccaggacaccaaag 34
          ||| |||||   ||  | || ||||||| |||||
Db    221 GCGGAAGCTGGAACGGTCAGACAGGACAGCAAAG 188


RESULT  10
US-08-612-973-5/c
; Sequence 5, Application US/08612973
; Patent No. 6150134
;  GENERAL INFORMATION:
;    APPLICANT:  MAERTENS, GEERT
;    APPLICANT:  BOSMAN, FONS
;    APPLICANT:  DE MARTYNOFF, GUY
;    APPLICANT:  BUYSE, MARIE-ANGE
;    TITLE OF INVENTION:  PURIFIED HEPATITIS C VIRUS ENVELOPE
;    TITLE OF INVENTION:  PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
;    NUMBER OF SEQUENCES:  111
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 NORTH GLEBE ROAD
;      CITY:  ARLINGTON
;      STATE:  VIRGINIA
;      COUNTRY:  U.S.A.
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/612,973
;      FILING DATE:  11-MAR-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BYRNE, THOMAS E.
;      REGISTRATION NUMBER:  32,205
;      REFERENCE/DOCKET NUMBER:  1487-10
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 816-4000
;      TELEFAX:  (703) 816-4100
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  795 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..792
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  1..789
US-08-612-973-5


  Query Match             54.4%;  Score 19.6;  DB 3;  Length 795;
  Best Local Similarity   73.5%;  Pred. No. 28;
  Matches   25;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      1 gcgaaagcttccaccatgagccaggacaccaaag 34
          ||| |||||   ||  | || ||||||| |||||
Db    221 GCGGAAGCTGGAACGGTCAGACAGGACAGCAAAG 188


RESULT  11
US-08-927-597-5/c
```

```
; Sequence 5, Application US/08927597
; Patent No. 6245503
;  GENERAL INFORMATION:
;    APPLICANT:  MAERTENS, GEERT
;    APPLICANT:  BOSMAN, FONS
;    APPLICANT:  DE MARTYNOFF, GUY
;    APPLICANT:  BUYSE, MARIE-ANGE
;    TITLE OF INVENTION:  PURIFIED HEPATITIS C VIRUS ENVELOPE
;    TITLE OF INVENTION:  PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
;    NUMBER OF SEQUENCES:  111
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 NORTH GLEBE ROAD
;      CITY:  ARLINGTON
;      STATE:  VIRGINIA
;      COUNTRY:  U.S.A.
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/927,597
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/612,973
;      FILING DATE:  11-MAR-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BYRNE, THOMAS E.
;      REGISTRATION NUMBER:  32,205
;      REFERENCE/DOCKET NUMBER:  1487-10
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 816-4000
;      TELEFAX:  (703) 816-4100
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  795 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..792
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  1..789
US-08-927-597-5


  Query Match             54.4%;  Score 19.6;  DB 4;  Length 795;
  Best Local Similarity   73.5%;  Pred. No. 28;
  Matches   25;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       1 gcgaaagcttccaccatgagccaggacaccaaag 34
           ||| |||||   ||  | || ||||||| |||||
Db     221 GCGGAAGCTGGAACGGTCAGACAGGACAGCAAAG 188


RESULT  12
US-08-612-973-47/c
; Sequence 47, Application US/08612973
; Patent No. 6150134
;  GENERAL INFORMATION:
;    APPLICANT:  MAERTENS, GEERT
;    APPLICANT:  BOSMAN, FONS
;    APPLICANT:  DE MARTYNOFF, GUY
;    APPLICANT:  BUYSE, MARIE-ANGE
;    TITLE OF INVENTION:  PURIFIED HEPATITIS C VIRUS ENVELOPE
;    TITLE OF INVENTION:  PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
;    NUMBER OF SEQUENCES:  111
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 NORTH GLEBE ROAD
;      CITY:  ARLINGTON
;      STATE:  VIRGINIA
;      COUNTRY:  U.S.A.
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/612,973
;      FILING DATE:  11-MAR-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BYRNE, THOMAS E.
;      REGISTRATION NUMBER:  32,205
;      REFERENCE/DOCKET NUMBER:  1487-10
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 816-4000
;      TELEFAX:  (703) 816-4100
;  INFORMATION FOR SEQ ID NO:  47:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2082 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..2079
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  1..2076
US-08-612-973-47


  Query Match             54.4%;  Score 19.6;  DB 3;  Length 2082;
  Best Local Similarity   73.5%;  Pred. No. 31;
  Matches   25;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       1 gcgaaagcttccaccatgagccaggacaccaaag 34
           ||| |||||   ||  | || ||||||| |||||
Db     221 GCGGAAGCTGGAACGGTCAGACAGGACAGCAAAG 188


RESULT  13
US-08-927-597-47/c
; Sequence 47, Application US/08927597
; Patent No. 6245503
;  GENERAL INFORMATION:
;    APPLICANT:  MAERTENS, GEERT
;    APPLICANT:  BOSMAN, FONS
;    APPLICANT:  DE MARTYNOFF, GUY
;    APPLICANT:  BUYSE, MARIE-ANGE
;    TITLE OF INVENTION:  PURIFIED HEPATITIS C VIRUS ENVELOPE
;    TITLE OF INVENTION:  PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
;    NUMBER OF SEQUENCES:  111
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 NORTH GLEBE ROAD
;      CITY:  ARLINGTON
;      STATE:  VIRGINIA
;      COUNTRY:  U.S.A.
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
```

```
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/927,597
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/612,973
;      FILING DATE:  11-MAR-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BYRNE, THOMAS E.
;      REGISTRATION NUMBER:  32,205
;      REFERENCE/DOCKET NUMBER:  1487-10
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 816-4000
;      TELEFAX:  (703) 816-4100
;  INFORMATION FOR SEQ ID NO:  47:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2082 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..2079
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  1..2076
US-08-927-597-47


  Query Match             54.4%;  Score 19.6;  DB 4;  Length 2082;
  Best Local Similarity   73.5%;  Pred. No. 31;
  Matches   25;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      1 gcgaaagcttccaccatgagccaggacaccaaag 34
          ||| |||||   ||  | || ||||||| |||||
Db    221 GCGGAAGCTGGAACGGTCAGACAGGACAGCAAAG 188


RESULT  14
US-08-612-973-49/c
; Sequence 49, Application US/08612973
; Patent No. 6150134
;  GENERAL INFORMATION:
;    APPLICANT:  MAERTENS, GEERT
;    APPLICANT:  BOSMAN, FONS
;    APPLICANT:  DE MARTYNOFF, GUY
;    APPLICANT:  BUYSE, MARIE-ANGE
;    TITLE OF INVENTION:  PURIFIED HEPATITIS C VIRUS ENVELOPE
;    TITLE OF INVENTION:  PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
;    NUMBER OF SEQUENCES:  111
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 NORTH GLEBE ROAD
;      CITY:  ARLINGTON
;      STATE:  VIRGINIA
;      COUNTRY:  U.S.A.
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/612,973
;      FILING DATE:  11-MAR-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BYRNE, THOMAS E.
;      REGISTRATION NUMBER:  32,205
;      REFERENCE/DOCKET NUMBER:  1487-10
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 816-4000
;      TELEFAX:  (703) 816-4100
;  INFORMATION FOR SEQ ID NO:  49:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2433 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..2430
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  1..2427
US-08-612-973-49


  Query Match             54.4%;  Score 19.6;  DB 3;  Length 2433;
  Best Local Similarity   73.5%;  Pred. No. 32;
  Matches   25;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      1 gcgaaagcttccaccatgagccaggacaccaaag 34
          ||| |||||   ||  | || ||||||| |||||
Db    572 GCGGAAGCTGGAACGGTCAGACAGGACAGCAAAG 539


RESULT  15
US-08-927-597-49/c
; Sequence 49, Application US/08927597
; Patent No. 6245503
;  GENERAL INFORMATION:
;    APPLICANT:  MAERTENS, GEERT
;    APPLICANT:  BOSMAN, FONS
;    APPLICANT:  DE MARTYNOFF, GUY
;    APPLICANT:  BUYSE, MARIE-ANGE
;    TITLE OF INVENTION:  PURIFIED HEPATITIS C VIRUS ENVELOPE
;    TITLE OF INVENTION:  PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
;    NUMBER OF SEQUENCES:  111
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 NORTH GLEBE ROAD
;      CITY:  ARLINGTON
;      STATE:  VIRGINIA
;      COUNTRY:  U.S.A.
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/927,597
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/612,973
;      FILING DATE:  11-MAR-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BYRNE, THOMAS E.
;      REGISTRATION NUMBER:  32,205
;      REFERENCE/DOCKET NUMBER:  1487-10
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 816-4000
;      TELEFAX:  (703) 816-4100
;  INFORMATION FOR SEQ ID NO:  49:
;    SEQUENCE CHARACTERISTICS:
```

; LENGTH: 2433 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2430
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 1..2427
US-08-927-597-49

Query Match 54.4%; Score 19.6; DB 4; Length 2433;
Best Local Similarity 73.5%; Pred. No. 32;
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```
Qy       1 gcgaaagcttccaccatgagccaggacaccaaag 34
           ||| ||||| || | || ||||||| ||||| 
Db     572 GCGGAAGCTGGAACGGTCAGACAGGACAGCAAAG 539
```

Search completed: June 20, 2002, 14:28:42
Job time: 18645 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2002, 12:16:58 ; Search time 4682.41 Seconds
(without alignments)
103.769 Million cell updates/sec

Title: US-09-927-267-12
Perfect score: 36
Sequence: 1 gcgaaagcttccaccatgagccaggacaccaaagtg 36

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	24.4	67.8	468	9	BB856662	BB856662 BB856662
c 2	23.8	66.1	102	9	AW793976	AW793976 MR1-UM000
c 3	23.8	66.1	1265	11	AK017058	AK017058 Mus muscu
4	23.6	65.6	181	9	AA020461	AA020461 mh61a08.r
5	23.6	65.6	258	9	BB863127	BB863127 BB863127
6	23.6	65.6	269	9	AA930603	AA930603 vy63d12.r
7	23.6	65.6	285	9	AA792349	AA792349 vs91e07.r
8	23.6	65.6	317	9	AA681601	AA681601 vr38e12.r
9	23.6	65.6	358	9	AI956505	AI956505 ul78b02.y
10	23.6	65.6	364	9	BB847706	BB847706 BB847706
11	23.6	65.6	386	9	AW229529	AW229529 uo60c12.y
12	23.6	65.6	400	9	AL363590	AL363590 AL363590
13	23.6	65.6	416	9	BB845746	BB845746 BB845746
14	23.6	65.6	418	9	AW211332	AW211332 uo79g02.y
15	23.6	65.6	423	9	AA981310	AA981310 vx60b01.r
16	23.6	65.6	431	9	BB864255	BB864255 BB864255
17	23.6	65.6	433	9	AA072327	AA072327 mm69c10.r
18	23.6	65.6	437	10	BF225057	BF225057 uz42d11.y
19	23.6	65.6	446	9	AI317054	AI317054 uk71h06.y
20	23.6	65.6	455	9	AA791484	AA791484 vs66g07.r
21	23.6	65.6	456	9	BB839570	BB839570 BB839570
22	23.6	65.6	460	9	BB860016	BB860016 BB860016
23	23.6	65.6	461	9	AI662984	AI662984 uj64c12.y
24	23.6	65.6	462	9	BB858988	BB858988 BB858988
25	23.6	65.6	465	9	BB861253	BB861253 BB861253
26	23.6	65.6	468	9	BB855989	BB855989 BB855989
27	23.6	65.6	474	9	BB851638	BB851638 BB851638
28	23.6	65.6	476	10	BF168440	BF168440 601776032
29	23.6	65.6	480	9	BB853977	BB853977 BB853977
30	23.6	65.6	480	9	AA492864	AA492864 vi79f06.r
31	23.6	65.6	496	9	BB853844	BB853844 BB853844
32	23.6	65.6	498	9	AA771051	AA771051 vt15h10.r
33	23.6	65.6	499	9	AI787389	AI787389 uj31c12.y
34	23.6	65.6	507	9	BB867656	BB867656 BB867656
35	23.6	65.6	515	9	AW761939	AW761939 ur51f05.y
36	23.6	65.6	521	9	AW762072	AW762072 ur53d01.y
37	23.6	65.6	527	10	BG146311	BG146311 mab90b07.
38	23.6	65.6	536	9	AA645307	AA645307 vs81b10.r
39	23.6	65.6	562	10	BE373844	BE373844 601227857
40	23.6	65.6	572	9	AW907769	AW907769 ur93e06.y
41	23.6	65.6	579	10	BG078049	BG078049 H3023A07-
42	23.6	65.6	579	10	BE381140	BE381140 601271491
43	23.6	65.6	589	10	BF301911	BF301911 602032939
44	23.6	65.6	594	10	BG873250	BG873250 602794490
45	23.6	65.6	595	9	AI227078	AI227078 uj11f04.y

ALIGNMENTS

RESULT 1
BB856662
LOCUS BB856662 468 bp mRNA linear EST 26-NOV-2001
DEFINITION BB856662 RIKEN full-length enriched, B16 F10Y cells Mus musculus cDNA clone G370036D04 5', mRNA sequence.
ACCESSION BB856662
VERSION BB856662.1 GI:17098116
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 468)
AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura

```
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES              Location/Qualifiers
     source           1. .468
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="G370036D04"
                      /clone_lib="RIKEN full-length enriched, B16 F10Y cells"
                      /cell_type="B16 F10Y cells"
BASE COUNT       96 a     116 c     176 g      80 t
ORIGIN


  Query Match             67.8%;  Score 24.4;  DB 9;  Length 468;
  Best Local Similarity   96.2%;  Pred. No. 33;
  Matches   25;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      11 ccaccatgagccaggacaccaaagtg 36
           |||||||||||||||||||| |||||
Db     124 CCACCATGAGCCAGGACACCGAAGTG 149


RESULT  2
AW793976/c
LOCUS       AW793976                 102 bp    mRNA    linear   EST 16-MAY-2000
DEFINITION  MR1-UM0009-210300-007-d02 UM0009 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW793976
VERSION     AW793976.1  GI:7845846
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 102)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=MR1-UM0009-210
            300-007-d02&t3=2000-03-21&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 10
            High quality sequence stop: 102.
FEATURES              Location/Qualifiers
     source           1. .102
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone_lib="UM0009"
                      /dev_stage="Adult"
                      /note="Organ: uterus; Vector: puc18; Site_1: SmaI; Site_2:
                      SmaI; A mini-library was made by cloning products derived
                      from  ORESTES PCR (U.S. Letters Patent application No. 196
                      ,716 -  Ludwig Institute for Cancer Research) profiles
                      into  the pUC 18 vector. Reverse transcription of tissue
                      mRNA and cDNA amplification were performed under low
                      stringency conditions."
BASE COUNT       19 a      30 c      26 g      27 t
ORIGIN


  Query Match             66.1%;  Score 23.8;  DB 9;  Length 102;
  Best Local Similarity   92.6%;  Pred. No. 33;
  Matches   25;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      10 tccaccatgagccaggacaccaaagtg 36
           |||||||||||||||||||| | |||
Db      89 TCCACCATGAGCCAGGACACCGAGGTG 63


RESULT  3
AK017058/c
LOCUS       AK017058                 1265 bp    mRNA    linear   HTC 19-JAN-2002
DEFINITION  Mus musculus adult male testis cDNA, RIKEN full-length enriched
            library, clone:4933434J16:testicular haploid expressed gene, full
            insert sequence.
ACCESSION   AK017058
VERSION     AK017058.1  GI:12856123
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:4933434J16.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1265)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
```

```
            Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
            Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
            Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
            Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
            Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
            Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
            Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
            Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
            Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I.,
            Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
            Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 10.0 and subtraction to
            Rot = 100.0. Second strand cDNA was prepared with the primer
            adapter of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved
            with BamHI and XhoI. cDNA of size comprised between 0.5 and 3 kb
            was selected before cloning. Vector: a modified pBluescript KS(+)
            after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI;
            3' end: BamHI. Host: DH10B.
FEATURES             Location/Qualifiers
     source          1. .1265
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="MGD:MGI:1903802"
                     /db_xref="taxon:10090"
                     /clone="4933434J16"
                     /sex="male"
                     /tissue_type="testis"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     gene            1. .1265
                     /gene="Theg"
     misc_feature    1. .1265
                     /gene="Theg"
                     /note="data source:MGD, source key:MGI:1338756,
                     evidence:ISS
                     testicular haploid expressed gene"
BASE COUNT      322 a    350 c    350 g    243 t
ORIGIN


  Query Match             66.1%;  Score 23.8;  DB 11;  Length 1265;
  Best Local Similarity   80.0%;  Pred. No. 79;
  Matches   28;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      2 cgaaagcttccaccatgagccaggacaccaaagtg 36
          |||||||| || ||||| ||  |||||||| || |
Db    558 CGAAAGCTCCCTCCATGCGCTGGGACACCACAGGG 524


RESULT   4
AA020461
LOCUS       AA020461                181 bp    mRNA    linear   EST 21-JAN-1997
DEFINITION  mh61a08.r1 Soares mouse placenta 4NbMP13.5 14.5 Mus musculus cDNA
            clone IMAGE:455414 5' similar to gb:J02939 4F2 CELL-SURFACE ANTIGEN
            HEAVY CHAIN (HUMAN); gb:X14309 Murine mRNA for 4F2 antigen heavy
            chain (MOUSE);, mRNA sequence.
ACCESSION   AA020461
VERSION     AA020461.1  GI:1484251
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 181)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:272302
            Trace considered overall poor quality
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .181
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:455414"
                     /clone_lib="Soares mouse placenta 4NbMP13.5 14.5"
                     /sex="unknown"
                     /tissue_type="placenta"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: placenta; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Not I - oligo(dT) primer
                     [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT       44 a     51 c     56 g     30 t
ORIGIN


  Query Match             65.6%;  Score 23.6;  DB 9;  Length 181;
  Best Local Similarity   86.7%;  Pred. No. 48;
  Matches   26;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      7 gcttccaccatgagccaggacaccaaagtg 36
            ||   |||||||||||||||||| |||||
Db     82 GCAGGCACCATGAGCCAGGACACCGAAGTG 111


RESULT   5
BB863127
LOCUS       BB863127                258 bp    mRNA    linear   EST 27-NOV-2001
DEFINITION  BB863127 RIKEN full-length enriched, pooled cell lines Mus musculus
            cDNA clone G430034G19 5', mRNA sequence.
ACCESSION   BB863127
VERSION     BB863127.1  GI:17109337
```

```
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 258)
  AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
            Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
            ,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
            Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.
            , Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
            Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
            ,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.
            , Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
            2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .258
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="G430034G19"
                     /clone_lib="RIKEN full-length enriched, pooled cell lines"
                     /note="pooled cell lines ; (cell_line=CRL-1751 WEHI 164),
                     (cell_line=CRL-2116 JC), (cell_line=RCB-0035 WEHI-3),
                     (cell_line=RCB-0464 Meth-A), (cell_line=RCB-0545 OHTA),
                     (cell_line=RCB-0559 K-1 . F1), (cell_line=RCB-1283 B16
                     melanoma), (cell_type=B cells, cell_line=CRL-1702 WEHI 231
                     ), (cell_type=Leydig cells, cell_line=CRL-2065 MLTC-1),
                     (cell_type=Nullipotent stem cell, cell_line=CRL-2070 NE),
                     (tissue_type=bladder, cell_line=RCB-0544 MBT-2),
                     (tissue_type=bone marrow, cell_type=stroma cell,
                     cell_line= CRL-2028 SR-4987), (tissue_type=colon,
                     cell_line=RCB-0549 Cle-H3), (tissue_type=kidney,
                     cell_line=CCL-142 RAG), (tissue_type=submandibular gland,
                     cell_line=CRL-1734 SCA-9 clone 15), (strain=BALB/C,
                     cell_type=B cells, cell_line=CRL-1669 BCL1 Clone 13.20-3B3
                     ), (strain=C3H, tissue_type=brain, cell_line=CRL-1443
                     BC3H1)"
BASE COUNT       65 a     56 c    102 g     35 t
ORIGIN


  Query Match             65.6%;  Score 23.6;  DB 9;  Length 258;
  Best Local Similarity   86.7%;  Pred. No. 55;
  Matches   26;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       7 gcttccaccatgagccaggacaccaaagtg 36
           ||   |||||||||||||||||| |||||
Db      86 GCAGGCACCATGAGCCAGGACACCGAAGTG 115


RESULT   6
AA930603
LOCUS       AA930603                 269 bp    mRNA    linear   EST 23-APR-1998
DEFINITION .vy63d12.r1 Stratagene mouse macrophage (#937306) Mus musculus cDNA
            clone IMAGE:1310903 5' similar to gb:J02939 4F2 CELL-SURFACE
            ANTIGEN HEAVY CHAIN (HUMAN); gb:X14309 Murine mRNA for 4F2 antigen
            heavy chain (MOUSE);, mRNA sequence.
ACCESSION   AA930603
VERSION     AA930603.1  GI:3079317
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 269)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:681199
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 38.
FEATURES             Location/Qualifiers
     source          1. .269
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1310903"
                     /clone_lib="Stratagene mouse macrophage (#937306)"
                     /tissue_type="macrophage"
                     /dev_stage="WEHI-3 cell line"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: blood; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. WEHI-3 cell line. Average insert size: 1.5 kb;
                     Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCACGAG
                     3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT       63 a     72 c     86 g     48 t
ORIGIN


  Query Match             65.6%;  Score 23.6;  DB 9;  Length 269;
  Best Local Similarity   86.7%;  Pred. No. 55;
  Matches   26;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       7 gcttccaccatgagccaggacaccaaagtg 36
           ||   |||||||||||||||||| |||||
Db      82 GCAGGCACCATGAGCCAGGACACCGAAGTG 111


RESULT   7
AA792349
LOCUS       AA792349                 285 bp    mRNA    linear   EST 09-FEB-1998
DEFINITION  vs91e07.r1 Barstead mouse myotubes MPLRB5 Mus musculus cDNA clone
```

```
            IMAGE:1153668 5' similar to gb:J02939 4F2 CELL-SURFACE ANTIGEN
            HEAVY CHAIN (HUMAN); gb:X14309 Murine mRNA for 4F2 antigen heavy
            chain (MOUSE);, mRNA sequence.
ACCESSION   AA792349
VERSION     AA792349.1  GI:2855304
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 285)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:626876
            Trace considered overall poor quality
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .285
                     /organism="Mus musculus"
                     /strain="C3H"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1153668"
                     /clone_lib="Barstead mouse myotubes MPLRB5"
                     /cell_line="C2C12"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     [AATTCGGATCCTTG], digested with Not I and cloned into the
                     Not I  and Eco RI sites of the modified pT7T3 vector.
                     Library constructed  by Bob Barstead.  The C2C12 cell line
                     (available from ATCC,  catalog # CRL-1772) differentiates
                     rapidly, forming contractile  myotubes and producing
                     characteristic muscle proteins.  "
BASE COUNT       69 a     66 c     96 g     54 t
ORIGIN


  Query Match             65.6%;  Score 23.6;  DB 9;  Length 285;
  Best Local Similarity   86.7%;  Pred. No. 56;
  Matches   26;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      7 gcttccaccatgagccaggacaccaaagtg 36
          ||   |||||||||||||||||| |||||
Db     83 GCAGGCACCATGAGCCAGGACACCGAAGTG 112


RESULT   8
AA681601
LOCUS       AA681601                317 bp    mRNA    linear   EST 05-DEC-1997
DEFINITION  vr38e12.r1 Barstead mouse myotubes MPLRB5 Mus musculus cDNA clone
            IMAGE:1122958 5' similar to gb:J02939 4F2 CELL-SURFACE ANTIGEN
            HEAVY CHAIN (HUMAN); gb:X14309 Murine mRNA for 4F2 antigen heavy
            chain (MOUSE);, mRNA sequence.
ACCESSION   AA681601
VERSION     AA681601.1  GI:2663741
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 317)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:612294
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 166.
FEATURES             Location/Qualifiers
     source          1. .317
                     /organism="Mus musculus"
                     /strain="C3H"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1122958"
                     /clone_lib="Barstead mouse myotubes MPLRB5"
                     /cell_line="C2C12"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     [AATTCGGATCCTTG], digested with Not I and cloned into the
                     Not I  and Eco RI sites of the modified pT7T3 vector.
                     Library constructed  by Bob Barstead.  The C2C12 cell line
                     (available from ATCC,  catalog # CRL-1772) differentiates
                     rapidly, forming contractile  myotubes and producing
                     characteristic muscle proteins.  "
BASE COUNT       77 a     75 c     112 g     52 t      1 others
ORIGIN


  Query Match             65.6%;  Score 23.6;  DB 9;  Length 317;
  Best Local Similarity   86.7%;  Pred. No. 59;
  Matches   26;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      7 gcttccaccatgagccaggacaccaaagtg 36
          ||   |||||||||||||||||| |||||
Db     76 GCAGGCACCATGAGCCAGGACACCGAAGTG 105


RESULT   9
AI956505
LOCUS       AI956505                358 bp    mRNA    linear   EST 20-AUG-1999
DEFINITION  ul78b02.y1 Sugano mouse kidney mkia Mus musculus cDNA clone
            IMAGE:2136651 5' similar to gb:J02939 4F2 CELL-SURFACE ANTIGEN
            HEAVY CHAIN (HUMAN); gb:X14309 Murine mRNA for 4F2 antigen heavy
            chain (MOUSE);, mRNA sequence.
ACCESSION   AI956505
VERSION     AI956505.1  GI:5749214
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 358)
```

```
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1001327
            Seq primer: custom primer used
            High quality sequence stop: 345.
FEATURES             Location/Qualifiers
     source          1. .358
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:2136651"
                     /clone_lib="Sugano mouse kidney mkia"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                     be used to isolate the cDNA insert.  Size selection was
                     performed to exclude fragments <1.5kb.  Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science).  Custom primers  for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."
BASE COUNT       79 a     84 c    133 g     62 t
ORIGIN


  Query Match             65.6%;  Score 23.6;  DB 9;  Length 358;
  Best Local Similarity   86.7%;  Pred. No. 61;
  Matches   26;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      7 gcttccaccatgagccaggacaccaaagtg 36
          ||   |||||||||||||||||| |||||
Db     80 GCAGGCACCATGAGCCAGGACACCGAAGTG 109


RESULT  10
BB847706
LOCUS       BB847706                364 bp    mRNA    linear   EST 26-NOV-2001
DEFINITION  BB847706 RIKEN full-length enriched, adult male kidney Mus musculus
            cDNA clone F530010N22 5', mRNA sequence.
ACCESSION   BB847706
VERSION     BB847706.1  GI:17086081
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 364)
  AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
            Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
            ,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
            Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.
            , Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
            Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
            ,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.
            , Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
            2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .364
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="F530010N22"
                     /clone_lib="RIKEN full-length enriched, adult male kidney"
                     /sex="male"
                     /tissue_type="kidney"
                     /dev_stage="adult"
                     /lab_host="SOLR"
                     /note="Site_1: XhoI; Site_2: SstI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCAATTAATTAATTAAACCCCCCCCCCC 3']. 
                     cDNA was cleaved with XhoI and SstI. "
BASE COUNT       84 a     93 c    133 g     54 t
ORIGIN


  Query Match             65.6%;  Score 23.6;  DB 9;  Length 364;
  Best Local Similarity   86.7%;  Pred. No. 61;
  Matches   26;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      7 gcttccaccatgagccaggacaccaaagtg 36
          ||   |||||||||||||||||| |||||
Db    118 GCAGGCACCATGAGCCAGGACACCGAAGTG 147


RESULT  11
AW229529
```

```
LOCUS       AW229529                386 bp    mRNA    linear   EST 10-DEC-1999
DEFINITION  uo60c12.y1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:2646934 5'
            similar to gb:J02939 4F2 CELL-SURFACE ANTIGEN HEAVY CHAIN (HUMAN);
            gb:X14309 Murine mRNA for 4F2 antigen heavy chain (MOUSE);, mRNA
            sequence.
ACCESSION   AW229529
VERSION     AW229529.1  GI:6558825
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 386)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Other_ESTs: uo60c12.x1
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            MGI:1027386
            Seq primer: -40RP from Gibco
            High quality sequence stop: 245.
FEATURES             Location/Qualifiers
     source          1. .386
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:2646934"
                     /clone_lib="NCI_CGAP_Mam1"
                     /tissue_type="tumor, biopsy sample"
                     /dev_stage="10 months, virgin"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT       77 a     99 c    145 g     64 t      1 others
ORIGIN


  Query Match             65.6%;  Score 23.6;  DB 9;  Length 386;
  Best Local Similarity   86.7%;  Pred. No. 63;
  Matches   26;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       7 gcttccaccatgagccaggacaccaaagtg 36
           ||   |||||||||||||||||| |||||
Db      67 GCAGGCACCATGAGCCAGGACACCGAAGTG 96


RESULT  12
AL363590
LOCUS       AL363590                400 bp    mRNA    linear   EST 04-AUG-2000
DEFINITION  AL363590 ICRFp 522 and 523 Mus musculus cDNA clone R9303B34 5',
            mRNA sequence.
ACCESSION   AL363590
VERSION     AL363590.1  GI:9693751
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 400)
  AUTHORS   Eickhoff,H., Schuchhardt,J., Ivanov,I., Meier-Ewert,S., O'Brien,J.,
            Malik,A., Tandon,N., Wolski,E., Rohlfs,E., Nyarsik,L., Reinhardt,R.
            , Nietfeld,W. and Lehrach,H.
  TITLE     Tissue gene expression analysis using arrayed normalized cDNA
            libraries
  JOURNAL   Genome Res. (2000) In press
COMMENT     Contact: MPIMG
            Abt.Lehrach
            Max Planck Institut Fuer Molekulare Genetik
            Ihnestrasse 73, Berlin, 14195 Germany
            The cDNA libraries ICRFp 522 and 523 were normalized with
            oligonucleotide fingerprinting, resulting in a unique subset of
            5376 cDNA clones.
FEATURES             Location/Qualifiers
     source          1. .400
                     /organism="Mus musculus"
                     /strain="Black 6"
                     /db_xref="taxon:10090"
                     /clone="R9303B34"
                     /clone_lib="ICRFp 522 and 523"
                     /tissue_type="embryo"
                     /dev_stage="9 and 12 pc embryo"
BASE COUNT       84 a    102 c    105 g     70 t     39 others
ORIGIN


  Query Match             65.6%;  Score 23.6;  DB 9;  Length 400;
  Best Local Similarity   86.7%;  Pred. No. 63;
  Matches   26;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       7 gcttccaccatgagccaggacaccaaagtg 36
           ||   |||||||||||||||||| |||||
Db     119 GCAGGCACCATGAGCCAGGACACCGAAGTG 148


RESULT  13
BB845746
LOCUS       BB845746                416 bp    mRNA    linear   EST 26-NOV-2001
DEFINITION  BB845746 RIKEN full-length enriched, adult male kidney Mus musculus
            cDNA clone F520007P09 5', mRNA sequence.
ACCESSION   BB845746
VERSION     BB845746.1  GI:17084121
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 416)
  AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
            Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
            ,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
            Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.
            , Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
            Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
            ,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.
            , Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
            2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
```

```
          ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
          Hayashizaki,Y.
           RIKEN integrated sequence analysis (RISA) system--384-format
          sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
          10 (11), 1757-1771 (2000)
           Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
          ,Y. and Hayashizaki,Y.
           Computer-based methods for the mouse full-length cDNA
          encyclopedia: real-time sequence clustering for construction of a
          nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
           Please visit our web site (http://genome.gsc.riken.go.jp) for
          further details.
           e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .416
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="F520007P09"
                     /clone_lib="RIKEN full-length enriched, adult male kidney"
                     /sex="male"
                     /tissue_type="kidney"
                     /dev_stage="adult"
                     /lab_host="SOLR"
                     /note="Site_1: XhoI; Site_2: SstI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCAATTAATTAATTAAACCCCCCCCCCC 3'].
                     cDNA was cleaved with XhoI and SstI. "
BASE COUNT       85 a    110 c    152 g     69 t
ORIGIN


  Query Match             65.6%;  Score 23.6;  DB 9;  Length 416;
  Best Local Similarity   86.7%;  Pred. No. 64;
  Matches   26;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       7 gcttccaccatgagccaggacaccaaagtg 36
           ||   |||||||||||||||||| |||||
Db     118 GCAGGCACCATGAGCCAGGACACCGAAGTG 147


RESULT  14
AW211332
LOCUS       AW211332                 418 bp    mRNA    linear   EST 03-DEC-1999
DEFINITION  uo79g02.y1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:2648786 5'
            similar to gb:J02939 4F2 CELL-SURFACE ANTIGEN HEAVY CHAIN (HUMAN);
            gb:X14309~Murine mRNA for 4F2 antigen heavy chain (MOUSE);, mRNA
            sequence.
ACCESSION   AW211332
VERSION     AW211332.1  GI:6517280
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 418)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Other_ESTs: uo79g02.x1
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            MGI:1029238
            Seq primer: -40RP from Gibco
            High quality sequence stop: 279.
FEATURES             Location/Qualifiers
     source          1. .418
                     /organism="Mus musculus"
                     /strain="129 - C57/B6 - FVBN"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:2648786"
                     /clone_lib="NCI_CGAP_Mam3"
                     /tissue_type="tumor, gross tissue"
                     /dev_stage="10 months"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigators
                     providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
                     Reference for transgenic model: Xu et al., Nature Genetics
                     22, 37-43 (1999)."
BASE COUNT       86 a    110 c    149 g     73 t
ORIGIN


  Query Match             65.6%;  Score 23.6;  DB 9;  Length 418;
  Best Local Similarity   86.7%;  Pred. No. 64;
  Matches   26;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       7 gcttccaccatgagccaggacaccaaagtg 36
           ||   |||||||||||||||||| |||||
Db     110 GCAGGCACCATGAGCCAGGACACCGAAGTG 139


RESULT  15
AA981310
LOCUS       AA981310                 423 bp    mRNA    linear   EST 27-MAY-1998
DEFINITION  vx60b01.r1 Stratagene mouse macrophage (#937306) Mus musculus cDNA
            clone IMAGE:1279561 5' similar to gb:X14309 Murine mRNA for 4F2
            antigen heavy chain (MOUSE);, mRNA sequence.
ACCESSION   AA981310
VERSION     AA981310.1  GI:3159846
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 423)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:671361
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 375.
FEATURES             Location/Qualifiers
```

```
     source          1. .423
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1279561"
                     /clone_lib="Stratagene mouse macrophage (#937306)"
                     /tissue_type="macrophage"
                     /dev_stage="WEHI-3 cell line"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: blood; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. WEHI-3 cell line. Average insert size: 1.5 kb;
                     Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCACGAG
                     3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT       86 a    108 c    159 g     70 t
ORIGIN


  Query Match            65.6%;  Score 23.6;  DB 9;  Length 423;
  Best Local Similarity  86.7%;  Pred. No. 65;
  Matches   26;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       7 gcttccaccatgagccaggacaccaaagtg 36
           ||   |||||||||||||||||| |||||
Db      86 GCAGGCACCATGAGCCAGGACACCGAAGTG 115


Search completed: June 20, 2002, 12:17:01
Job time: 11374 sec
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2002, 14:35:35 ; Search time 7623.77 Seconds
(without alignments)
96.072 Million cell updates/sec

Title: US-09-927-267-13
Perfect score: 35
Sequence: 1 catgtctagaatggggatggggtcactctggacct 35

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	25.4	72.6	1835	6	AX350402	AX350402 Sequence
c 2	25.4	72.6	2013	6	AX355967	AX355967 Sequence
c 3	25.4	72.6	2551	6	AX350404	AX350404 Sequence
c 4	25.4	72.6	162965	2	AC021935	AC021935 Homo sapi
c 5	25.4	72.6	164434	2	AC036216	AC036216 Homo sapi
c 6	25.4	72.6	186609	2	AC022762	AC022762 Homo sapi
c 7	23	65.7	50362	2	AC091572	AC091572 Homo sapi
8	23	65.7	94162	9	AL450322	AL450322 Human DNA
9	23	65.7	143946	2	AL607063	AL607063 Homo sapi
c 10	23	65.7	169060	2	AC023903	AC023903 Homo sapi
11	22.4	64.0	98951	2	AC090069	AC090069 Homo sapi
c 12	22.4	64.0	208197	9	AC009107	AC009107 Homo sapi
13	21.8	62.3	3536	10	MMU44942	U44942 Mus musculu
14	21.8	62.3	139182	9	AL137851	AL137851 Human DNA
15	21.8	62.3	166169	9	HS527D15	AL513422 Homo sapi
16	21.4	61.1	1532	8	AF168776	AF168776 Eucalyptu
c 17	21.4	61.1	53336	9	AC092600	AC092600 Homo sapi
c 18	21.4	61.1	58554	2	AC100623	AC100623 Mus muscu
19	21.4	61.1	125290	9	AC004986	AC004986 Homo sapi
c 20	21.4	61.1	137830	2	AC024346	AC024346 Homo sapi
21	21.4	61.1	142545	2	AL590610	AL590610 Homo sapi
c 22	21.4	61.1	170796	2	AC069416	AC069416 Homo sapi
23	21.4	61.1	196734	2	AC010640	AC010640 Homo sapi
c 24	21.4	61.1	198942	9	AC008517	AC008517 Homo sapi
25	21.2	60.6	57316	2	AC102779	AC102779 Mus muscu
26	21.2	60.6	157067	9	AC010163	AC010163 Homo sapi
c 27	21.2	60.6	195702	2	AC069537	AC069537 Homo sapi
28	21.2	60.6	206393	2	AC093450	AC093450 Mus muscu
29	21	60.0	54233	2	AC100456	AC100456 Mus muscu
30	21	60.0	59866	2	AC100572	AC100572 Mus muscu
c 31	21	60.0	122321	9	AC011472	AC011472 Homo sapi
c 32	21	60.0	173169	2	AC094464	AC094464 Rattus no
33	21	60.0	194144	2	AC095377	AC095377 Rattus no
34	20.8	59.4	140795	2	AC079562	AC079562 Mus muscu
35	20.8	59.4	145655	8	AP003546	AP003546 Oryza sat
c 36	20.8	59.4	153078	9	AC022493	AC022493 Homo sapi
c 37	20.8	59.4	163304	2	AC104467	AC104467 Homo sapi
c 38	20.8	59.4	169072	9	AC009946	AC009946 Homo sapi
39	20.8	59.4	173509	2	AC008557	AC008557 Homo sapi
40	20.8	59.4	184693	2	AC098195	AC098195 Rattus no
c 41	20.8	59.4	185804	2	AC098181	AC098181 Rattus no
42	20.8	59.4	190487	2	AC084310	AC084310 Mus muscu
c 43	20.8	59.4	191804	2	AC019017	AC019017 Homo sapi
c 44	20.8	59.4	197025	2	AC105642	AC105642 Rattus no
45	20.8	59.4	218687	2	AC079501	AC079501 Mus muscu

ALIGNMENTS

RESULT 1
AX350402/c
LOCUS AX350402 1835 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 27 from Patent WO0181578.
ACCESSION AX350402
VERSION AX350402.1 GI:18616045
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Vernet,C.A., Fernandes,E.R., Gerlach,V., Shimkets,R.A.,
Malyankar,U.M., Boldog,F.L., Zerhusen,B.D., Spytek,K.A.,
Majumder,K., Tchernev,V.T., Padigaru,M., Patturajan,M.,
Burgess,C.E., Gangolli,E.A., Smithson,G., Rastelli,L.,
Macdougall,J.R., Taupier,R.J., Grosse,W.M. and Alsobrook,J.P.
TITLE Novel proteins and nucleic acids encoding same
JOURNAL Patent: WO 0181578-A 27 01-NOV-2001;
Curagen Corporation (US)
FEATURES Location/Qualifiers
source 1. .1835
/organism="Homo sapiens"

```
                     /db_xref="taxon:9606"
BASE COUNT      420 a    521 c    521 g    373 t
ORIGIN


  Query Match            72.6%;  Score 25.4;  DB 6;  Length 1835;
  Best Local Similarity  96.3%;  Pred. No. 4.5;
  Matches   26;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      9 gaatggggatggggtcactctggacct 35
          | ||||||||||||||||||||||||||
Db   1804 GGATGGGGATGGGGTCACTCTGGACCT 1778


RESULT   2
AX355967/c
LOCUS       AX355967               2013 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 2 from Patent WO0202633.
ACCESSION   AX355967
VERSION     AX355967.1  GI:18620561
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Raumann,B.E., Sanjanwala,M.S., Tribouley,C.M. and Walia,N.K.
  TITLE     Transporters and ion channels
  JOURNAL   Patent: WO 0202633-A 2 10-JAN-2002;
            Incyte Genomics, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .2013
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /note="Incyte ID No: 6703242CB1"
BASE COUNT      472 a    576 c    557 g    408 t
ORIGIN


  Query Match            72.6%;  Score 25.4;  DB 6;  Length 2013;
  Best Local Similarity  96.3%;  Pred. No. 4.5;
  Matches   26;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      9 gaatggggatggggtcactctggacct 35
          | ||||||||||||||||||||||||||
Db   1865 GGATGGGGATGGGGTCACTCTGGACCT 1839


RESULT   3
AX350404/c
LOCUS       AX350404               2551 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 29 from Patent WO0181578.
ACCESSION   AX350404
VERSION     AX350404.1  GI:18616046
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Vernet,C.A., Fernandes,E.R., Gerlach,V., Shimkets,R.A.,
            Malyankar,U.M., Boldog,F.L., Zerhusen,B.D., Spytek,K.A.,
            Majumder,K., Tchernev,V.T., Padigaru,M., Patturajan,M.,
            Burgess,C.E., Gangolli,E.A., Smithson,G., Rastelli,L.,
            Macdougall,J.R., Taupier,R.J., Grosse,W.M. and Alsobrook,J.P.
  TITLE     Novel proteins and nucleic acids encoding same
  JOURNAL   Patent: WO 0181578-A 29 01-NOV-2001;
            Curagen Corporation (US)
FEATURES             Location/Qualifiers
     source          1. .2551
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      646 a    668 c    707 g    530 t
ORIGIN


  Query Match            72.6%;  Score 25.4;  DB 6;  Length 2551;
  Best Local Similarity  96.3%;  Pred. No. 4.3;
  Matches   26;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      9 gaatggggatggggtcactctggacct 35
          | ||||||||||||||||||||||||||
Db   2520 GGATGGGGATGGGGTCACTCTGGACCT 2494


RESULT   4
AC021935/c
LOCUS       AC021935               162965 bp    DNA     linear   HTG 03-MAR-2001
DEFINITION  Homo sapiens chromosome 11 clone RP11-645I8 map 11, WORKING DRAFT
            SEQUENCE, 5 unordered pieces.
ACCESSION   AC021935
VERSION     AC021935.3  GI:13184160
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 162965)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-645I8
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 162965)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 2, 2001 this sequence version replaced gi:7652890.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5277
                Center clone name: 645_I_8
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 40% of reads
                Sequencing vector: Plasmid; n/a; 60% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 161664 bases at least Q40
                Consensus quality: 161989 bases at least Q30
                Consensus quality: 162229 bases at least Q20
                Insert size: 158000; agarose-fp
                Insert size: 162565; sum-of-contigs
                Quality coverage: 12.2 in Q20 bases; agarose-fp
                Quality coverage: 11.9 in Q20.
```

```
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     5772: contig of 5772 bp in length
            *     5773 5872: gap of      100 bp
            *     5873   114225: contig of 108353 bp in length
            *   114226 114325: gap of      100 bp
            *   114326   119496: contig of 5171 bp in length
            *   119497 119596: gap of      100 bp
            *   119597   161781: contig of 42185 bp in length
            *   161782 161881: gap of      100 bp
            *   161882   162965: contig of 1084 bp in length.
FEATURES             Location/Qualifiers
     source          1. .162965
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-645I8"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .5772
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    5873. .114225
                     /note="assembly_fragment"
     misc_feature    114326. .119496
                     /note="assembly_fragment"
     misc_feature    119597. .161781
                     /note="assembly_fragment"
     misc_feature    161882. .162965
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    54235 a  33886 c  32984 g  41460 t    400 others
ORIGIN


  Query Match             72.6%;  Score 25.4;  DB 2;  Length 162965;
  Best Local Similarity   96.3%;  Pred. No. 2.4;
  Matches   26;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      9 gaatggggatggggtcactctggacct 35
          | ||||||||||||||||||||||||||
Db 154883 GGATGGGGATGGGGTCACTCTGGACCT 154857


RESULT   5
AC036216/c
LOCUS       AC036216              164434 bp    DNA     linear   HTG 04-MAY-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-658A5 map 11, WORKING DRAFT
            SEQUENCE, 26 unordered pieces.
ACCESSION   AC036216
VERSION     AC036216.2  GI:7705203
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 164434)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-658A5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 164434)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 4, 2000 this sequence version replaced gi:7523886.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L9404
                Center clone name: 658_A_5
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 149016 bases at least Q40
                Consensus quality: 157301 bases at least Q30
                Consensus quality: 160298 bases at least Q20
                Insert size: 157000; agarose-fp
                Insert size: 161934; sum-of-contigs
                Quality coverage: 3.7 in Q20 bases; agarose-fp
                Quality coverage: 3.6 in Q20 bases; sum-of-contigs
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1416: contig of 1416 bp in length
            *     1417 1516: gap of      100 bp
            *     1517     3580: contig of 2064 bp in length
            *     3581 3680: gap of      100 bp
            *     3681     4383: contig of 703 bp in length
            *     4384 4483: gap of      100 bp
            *     4484     5844: contig of 1361 bp in length
            *     5845 5944: gap of      100 bp
            *     5945     7574: contig of 1630 bp in length
            *     7575 7674: gap of      100 bp
            *     7675     9988: contig of 2314 bp in length
            *     9989 10088: gap of      100 bp
            *    10089    13426: contig of 3338 bp in length
            *    13427 13526: gap of      100 bp
            *    13527    15573: contig of 2047 bp in length
            *    15574 15673: gap of      100 bp
            *    15674    18974: contig of 3301 bp in length
            *    18975 19074: gap of      100 bp
            *    19075    21809: contig of 2735 bp in length
            *    21810 21909: gap of      100 bp
            *    21910    25085: contig of 3176 bp in length
```

```
            *    25086 25185: gap of      100 bp
            *    25186    29425: contig of 4240 bp in length
            *    29426 29525: gap of      100 bp
            *    29526    34133: contig of 4608 bp in length
            *    34134 34233: gap of      100 bp
            *    34234    40649: contig of 6416 bp in length
            *    40650 40749: gap of      100 bp
            *    40750    45891: contig of 5142 bp in length
            *    45892 45991: gap of      100 bp
            *    45992    53180: contig of 7189 bp in length
            *    53181 53280: gap of      100 bp
            *    53281    62817: contig of 9537 bp in length
            *    62818 62917: gap of      100 bp
            *    62918    70247: contig of 7330 bp in length
            *    70248 70347: gap of      100 bp
            *    70348    78754: contig of 8407 bp in length
            *    78755 78854: gap of      100 bp
            *    78855    86840: contig of 7986 bp in length
            *    86841 86940: gap of      100 bp
            *    86941    98610: contig of 11670 bp in length
            *    98611 98710: gap of      100 bp
            *    98711   111069: contig of 12359 bp in length
            *   111070 111169: gap of      100 bp
            *   111170   122730: contig of 11561 bp in length
            *   122731 122830: gap of      100 bp
            *   122831   135722: contig of 12892 bp in length
            *   135723 135822: gap of      100 bp
            *   135823   150963: contig of 15141 bp in length
            *   150964 151063: gap of      100 bp
            *   151064   164434: contig of 13371 bp in length.
FEATURES             Location/Qualifiers
     source          1. .164434
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-658A5"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1416
                     /note="assembly_fragment"
     misc_feature    1517. .3580
                     /note="assembly_fragment"
     misc_feature    3681. .4383
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    4484. .5844
                     /note="assembly_fragment"
     misc_feature    5945. .7574
                     /note="assembly_fragment"
     misc_feature    7675. .9988
                     /note="assembly_fragment"
     misc_feature    10089. .13426
                     /note="assembly_fragment"
     misc_feature    13527. .15573
                     /note="assembly_fragment"
     misc_feature    15674. .18974
                     /note="assembly_fragment"
     misc_feature    19075. .21809
                     /note="assembly_fragment"
     misc_feature    21910. .25085
                     /note="assembly_fragment"
     misc_feature    25186. .29425
                     /note="assembly_fragment"
     misc_feature    29526. .34133
                     /note="assembly_fragment"
     misc_feature    34234. .40649
                     /note="assembly_fragment"
     misc_feature    40750. .45891
                     /note="assembly_fragment"
     misc_feature    45992. .53180
                     /note="assembly_fragment"
     misc_feature    53281. .62817
                     /note="assembly_fragment"
     misc_feature    62918. .70247
                     /note="assembly_fragment"
     misc_feature    70348. .78754
                     /note="assembly_fragment"
     misc_feature    78855. .86840
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    86941. .98610
                     /note="assembly_fragment"
     misc_feature    98711. .111069
                     /note="assembly_fragment"
     misc_feature    111170. .122730
                     /note="assembly_fragment"
     misc_feature    122831. .135722
                     /note="assembly_fragment"
     misc_feature    135823. .150963
                     /note="assembly_fragment"
     misc_feature    151064. .164434
                     /note="assembly_fragment"
BASE COUNT    50098 a  34031 c  34547 g  43240 t   2518 others
ORIGIN


  Query Match             72.6%;  Score 25.4;  DB 2;  Length 164434;
  Best Local Similarity   96.3%;  Pred. No. 2.4;
  Matches   26;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      9 gaatggggatggggtcactctggacct 35
          | ||||||||||||||||||||||||||
Db 161856 GGATGGGGATGGGGTCACTCTGGACCT 161830


RESULT   6
AC022762/c
LOCUS       AC022762              186609 bp    DNA     linear   HTG 09-FEB-2002
DEFINITION  Homo sapiens chromosome 11 clone RP11-290F24 map 11, *** SEQUENCING
            IN PROGRESS ***, 1 ordered pieces.
ACCESSION   AC022762
VERSION     AC022762.6  GI:18642872
KEYWORDS    HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 186609)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-290F24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 186609)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 9, 2002 this sequence version replaced gi:18483561.
            All repeats were identified using RepeatMasker:
```

```
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5842
                Center clone name: 290_F_24
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   186609: contig of 186609 bp in length.
FEATURES             Location/Qualifiers
     source          1. .186609
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-290F24"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    60812 a  39184 c  37829 g  48784 t
ORIGIN


  Query Match             72.6%;  Score 25.4;  DB 2;  Length 186609;
  Best Local Similarity   96.3%;  Pred. No. 2.4;
  Matches   26;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      9 gaatggggatggggtcactctggacct 35
          | ||||||||||||||||||||||||||
Db 160662 GGATGGGGATGGGGTCACTCTGGACCT 160636


RESULT   7
AC091572/c
LOCUS       AC091572               50362 bp    DNA     linear   HTG 28-NOV-2001
DEFINITION  Homo sapiens chromosome 8 clone RP11-184L12 map 8, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC091572
VERSION     AC091572.2  GI:17047138
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 50362)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-184L12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 50362)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
            Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K.,
            Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
            Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAY-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 22, 2001 this sequence version replaced gi:13959178.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L12175
                Center clone name: 184_L_12
            --------------.
            * NOTE: This record contains 66 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *       1      677: contig of 677 bp in length
            *     678 777: gap of      100 bp
            *     778     1461: contig of 684 bp in length
            *    1462 1561: gap of      100 bp
            *    1562     2228: contig of 667 bp in length
            *    2229 2328: gap of      100 bp
            *    2329     3004: contig of 676 bp in length
            *    3005 3104: gap of      100 bp
            *    3105     3785: contig of 681 bp in length
            *    3786 3885: gap of      100 bp
            *    3886     4534: contig of 649 bp in length
            *    4535 4634: gap of      100 bp
            *    4635     5305: contig of 671 bp in length
            *    5306 5405: gap of      100 bp
            *    5406     6054: contig of 649 bp in length
            *    6055 6154: gap of      100 bp
            *    6155     6817: contig of 663 bp in length
            *    6818 6917: gap of      100 bp
            *    6918     7568: contig of 651 bp in length
            *    7569 7668: gap of      100 bp
            *    7669     8344: contig of 676 bp in length
            *    8345 8444: gap of      100 bp
            *    8445     9106: contig of 662 bp in length
            *    9107 9206: gap of      100 bp
            *    9207     9878: contig of 672 bp in length
            *    9879 9978: gap of      100 bp
            *    9979    10658: contig of 680 bp in length
            *   10659 10758: gap of      100 bp
            *   10759    11430: contig of 672 bp in length
            *   11431 11530: gap of      100 bp
            *   11531    12161: contig of 631 bp in length
            *   12162 12261: gap of      100 bp
            *   12262    12925: contig of 664 bp in length
            *   12926 13025: gap of      100 bp
            *   13026    13689: contig of 664 bp in length
            *   13690 13789: gap of      100 bp
            *   13790    14464: contig of 675 bp in length
            *   14465 14564: gap of      100 bp
            *   14565    15224: contig of 660 bp in length
```

```
*    15225 15324: gap of      100 bp
*    15325    15990: contig of 666 bp in length
*    15991 16090: gap of      100 bp
*    16091    16755: contig of 665 bp in length
*    16756 16855: gap of      100 bp
*    16856    17517: contig of 662 bp in length
*    17518 17617: gap of      100 bp
*    17618    18287: contig of 670 bp in length
*    18288 18387: gap of      100 bp
*    18388    19056: contig of 669 bp in length
*    19057 19156: gap of      100 bp
*    19157    19835: contig of 679 bp in length
*    19836 19935: gap of      100 bp
*    19936    20602: contig of 667 bp in length
*    20603 20702: gap of      100 bp
*    20703    21363: contig of 661 bp in length
*    21364 21463: gap of      100 bp
*    21464    22142: contig of 679 bp in length
*    22143 22242: gap of      100 bp
*    22243    22884: contig of 642 bp in length
*    22885 22984: gap of      100 bp
*    22985    23650: contig of 666 bp in length
*    23651 23750: gap of      100 bp
*    23751    24405: contig of 655 bp in length
*    24406 24505: gap of      100 bp
*    24506    25185: contig of 680 bp in length
*    25186 25285: gap of      100 bp
*    25286    25958: contig of 673 bp in length
*    25959 26058: gap of      100 bp
*    26059    26729: contig of 671 bp in length
*    26730 26829: gap of      100 bp
*    26830    27492: contig of 663 bp in length
*    27493 27592: gap of      100 bp
*    27593    28275: contig of 683 bp in length
*    28276 28375: gap of      100 bp
*    28376    29032: contig of 657 bp in length
*    29033 29132: gap of      100 bp
*    29133    29794: contig of 662 bp in length
*    29795 29894: gap of      100 bp
*    29895    30573: contig of 679 bp in length
*    30574 30673: gap of      100 bp
*    30674    31343: contig of 670 bp in length
*    31344 31443: gap of      100 bp
*    31444    32106: contig of 663 bp in length
*    32107 32206: gap of      100 bp
*    32207    32865: contig of 659 bp in length
*    32866 32965: gap of      100 bp
*    32966    33628: contig of 663 bp in length
*    33629 33728: gap of      100 bp
*    33729    34400: contig of 672 bp in length
*    34401 34500: gap of      100 bp
*    34501    35160: contig of 660 bp in length
*    35161 35260: gap of      100 bp
*    35261    35916: contig of 656 bp in length
*    35917 36016: gap of      100 bp
*    36017    36673: contig of 657 bp in length
*    36674 36773: gap of      100 bp
*    36774    37432: contig of 659 bp in length
*    37433 37532: gap of      100 bp
*    37533    38195: contig of 663 bp in length
*    38196 38295: gap of      100 bp
*    38296    38972: contig of 677 bp in length
*    38973 39072: gap of      100 bp
*    39073    39747: contig of 675 bp in length
*    39748 39847: gap of      100 bp
*    39848    40515: contig of 668 bp in length
*    40516 40615: gap of      100 bp
*    40616    41260: contig of 645 bp in length
*    41261 41360: gap of      100 bp
*    41361    42025: contig of 665 bp in length
*    42026 42125: gap of      100 bp
*    42126    42811: contig of 686 bp in length
*    42812 42911: gap of      100 bp
*    42912    43571: contig of 660 bp in length
*    43572 43671: gap of      100 bp
*    43672    44324: contig of 653 bp in length
*    44325 44424: gap of      100 bp
*    44425    45091: contig of 667 bp in length
*    45092 45191: gap of      100 bp
*    45192    45854: contig of 663 bp in length
*    45855 45954: gap of      100 bp
*    45955    46632: contig of 678 bp in length
*    46633 46732: gap of      100 bp
*    46733    47338: contig of 606 bp in length
*    47339 47438: gap of      100 bp
*    47439    48068: contig of 630 bp in length
*    48069 48168: gap of      100 bp
*    48169    48846: contig of 678 bp in length
*    48847 48946: gap of      100 bp
*    48947    49606: contig of 660 bp in length
*    49607 49706: gap of      100 bp
*    49707    50362: contig of 656 bp in length.
FEATURES             Location/Qualifiers
     source          1. .50362
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"

  Query Match             65.7%;  Score 23;  DB 2;  Length 50362;
  Best Local Similarity   83.9%;  Pred. No. 28;
  Matches   26;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       3 tgtctagaatggggatggggtcactctggac 33
           |||| |||||||||||  ||| || |||||||
Db   10991 TGTCCAGAATGGGGAGCGGGCCATTCTGGAC 10961


RESULT   8
AL450322
LOCUS       AL450322               94162 bp    DNA     linear   PRI 05-OCT-2001
DEFINITION  Human DNA sequence from clone RP11-184A2 on chromosome 10, complete
            sequence.
ACCESSION   AL450322
VERSION     AL450322.17  GI:15983904
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 94162)
  AUTHORS   Kimberley,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 8, 2001 this sequence version replaced gi:14787539.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
```

```
            chromosome 10, constructed by the Sanger Centre Chromosome 10
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr10
            RP11-184A2 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-184A2 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true left end of clone RP11-184A2 is at 1 in this sequence. The
            true left end of clone RP11-464C19 is at 92163 in this sequence.
FEATURES              Location/Qualifiers
     source           1. .94162
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="10"
                      /clone="RP11-184A2"
                      /clone_lib="RPCI-11.1"
BASE COUNT    27141 a  20291 c  20119 g  26611 t
ORIGIN


  Query Match             65.7%;  Score 23;  DB 9;  Length 94162;
  Best Local Similarity   83.9%;  Pred. No. 25;
  Matches   26;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      3 tgtctagaatggggatggggtcactctggac 33
          |||| |||||||||||  ||| || |||||||
Db  33216 TGTCCAGAATGGGGAGCGGGCCATTCTGGAC 33246


RESULT   9
AL607063
LOCUS       AL607063              143946 bp    DNA     linear   HTG 17-NOV-2001
DEFINITION  Homo sapiens chromosome 10 clone RP11-409P6, *** SEQUENCING IN
            PROGRESS ***, in ordered pieces.
ACCESSION   AL607063
VERSION     AL607063.13  GI:17017819
KEYWORDS    HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Brown,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 20, 2001 this sequence version replaced gi:16974011.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA409P6
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 143897 bases at least Q40
            Consensus quality: 143923 bases at least Q30
            Consensus quality: 143936 bases at least Q20
            Insert size: 143946; sum-of-contigs
            Insert size: 150167; 1.5% error; agarose-fp
            Quality coverage: 11.02x in Q20 bases; sum-of-contigs Quality
            coverage: 10.66x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence..
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES              Location/Qualifiers
     source           1. .143946
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="10"
                      /clone="RP11-409P6"
                      /clone_lib="RPCI-11.2"
     misc_feature     1. .143946
                      /note="assembly_fragment:01299"
BASE COUNT    41390 a  31148 c  30897 g  40511 t
ORIGIN


  Query Match             65.7%;  Score 23;  DB 2;  Length 143946;
  Best Local Similarity   83.9%;  Pred. No. 24;
  Matches   26;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      3 tgtctagaatggggatggggtcactctggac 33
          |||| |||||||||||  ||| || |||||||
Db  97476 TGTCCAGAATGGGGAGCGGGCCATTCTGGAC 97506


RESULT  10
AC023903/c
LOCUS       AC023903              169060 bp    DNA     linear   HTG 30-AUG-2001
DEFINITION  Homo sapiens chromosome 10 clone RP11-237H5, WORKING DRAFT
            SEQUENCE, 4 unordered pieces.
ACCESSION   AC023903
VERSION     AC023903.8  GI:14787170
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN;
            HTGS_CANCELLED.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169060)
  AUTHORS   Smith,D.R.
  TITLE     Genome Therapeutics Corporation Sequencing Center: Human Genome
            Sequence Data
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169060)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-FEB-2000) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
COMMENT     On Jul 17, 2001 this sequence version replaced gi:14277221.
            -------------- Genome Center
                           Center: Genome Therapeutics Corperation
                           Center code: GTC
                           Web site: http://www.genomecorp.com/
                           Contact: gtc-seqcenter@genomecorp.com
                        -------------- Project Information
                           Center project name: hg192
                        -------------- Summary Statistics
                           Sequencing vector: N/A
                           Chemistry: Dye-terminator Big Dye; 100% of reads
                           Assembly program: Phrap; version 990315
                           Consensus quality: 164113 bases at least Q40
                           Consensus quality: 165662 bases at least Q30
                           Consensus quality: 166404 bases at least Q20
                           Insert size: 168760; sum-of-contigs
                           Quality coverage: 6.3x in Q20 bases; sum-of-contigs
                        --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1821: contig of 1821 bp in length
```

```
            *     1822     1921: gap of unknown length
            *     1922     3531: contig of 1610 bp in length
            *     3532     3631: gap of unknown length
            *     3632    36181: contig of 32550 bp in length
            *    36182    36281: gap of unknown length
            *    36282   169060: contig of 132779 bp in length.
FEATURES             Location/Qualifiers
     source          1. .169060
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-237H5"
                     /clone_lib="RPCI-11"
     misc_feature    1. .1821
                     /note="assembly_name:Contig2"
     misc_feature    1922. .3531
                     /note="assembly_name:Contig3"
     misc_feature    3632. .36181
                     /note="assembly_name:Contig4
                     clone_end:T7"
     misc_feature    36282. .169060
                     /note="assembly_name:Contig5
                     clone_end:SP6"
BASE COUNT    46703 a  36755 c  36892 g  48402 t    308 others
ORIGIN


  Query Match             65.7%;  Score 23;  DB 2;  Length 169060;
  Best Local Similarity   83.9%;  Pred. No. 23;
  Matches   26;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      3 tgtctagaatggggatggggtcactctggac 33
          |||| |||||||||||  ||| || |||||||
Db 121009 TGTCCAGAATGGGGAGCGGGCCATTCTGGAC 120979


RESULT  11
AC090069
LOCUS       AC090069               98951 bp    DNA     linear   HTG 20-APR-2001
DEFINITION  Homo sapiens chromosome 5 clone CTD-2162G15, WORKING DRAFT
            SEQUENCE, 7 unordered pieces.
ACCESSION   AC090069
VERSION     AC090069.1  GI:12830173
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 98951)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 98951)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-FEB-2001) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 689916
            Center clone name: CITB-H1_2162G15
            -------
            Summary Statistics
            Consensus quality: 91863 bases at least Q40
            Consensus quality: 96729 bases at least Q30
            Consensus quality: 97483 bases at least Q20
            Estimated insert size: 103000; agarose-fp estimation
            Estimated insert size: 98351; sum-of-contigs estimation
            Quality coverage:  5.7 in Q20 bases; agarose-fp estimation
            Quality coverage:  5.97 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1640: contig of 1640 bp in length
            *     1641     1740: gap of unknown length
            *     1741     4817: contig of 3077 bp in length
            *     4818     4917: gap of unknown length
            *     4918    10479: contig of 5562 bp in length
            *    10480    10579: gap of unknown length
            *    10580    17185: contig of 6606 bp in length
            *    17186    17285: gap of unknown length
            *    17286    27136: contig of 9851 bp in length
            *    27137    27236: gap of unknown length
            *    27237    53346: contig of 26110 bp in length
            *    53347    53446: gap of unknown length
            *    53447    98951: contig of 45505 bp in length.
FEATURES             Location/Qualifiers
     source          1. .98951
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-2162G15"
                     /clone_lib="CalTech human BAC library D"
BASE COUNT    28584 a  22060 c  20767 g  26935 t    605 others
ORIGIN


  Query Match             64.0%;  Score 22.4;  DB 2;  Length 98951;
  Best Local Similarity   81.2%;  Pred. No. 45;
  Matches   26;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 catgtctagaatggggatggggtcactctgga 32
            ||| | || |||||||||| |||  |||||||
Db  89341 CATCTGTAAAATGGGGATGTGGTGCCTCTGGA 89372


RESULT  12
AC009107/c
LOCUS       AC009107               208197 bp    DNA     linear   PRI 04-DEC-2001
DEFINITION  Homo sapiens chromosome 16 clone RP11-459F6, complete sequence.
ACCESSION   AC009107
VERSION     AC009107.10  GI:17298588
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 208197)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 208197)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 208197)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2001) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Dec 4, 2001 this sequence version replaced gi:13786305.
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
```

```
            1 error in 10,000 bp.
            Base-by-base quality values are not generally visible from the
            GenBank flat file format but are available as part
            of this entry's ASN.1 file.
            ----------------------
            Sequence Quality Assessment:
            This entry has been annotated with sequence quality
            estimates computed by the Phrap assembly program.
            All manually edited bases have been reduced to quality zero.
            Quality levels above 40 are expected to have less than
            1 error in 10,000 bp.
            Base-by-base quality values are not generally visible from the
            GenBank flat file format but are available as part
            of this entry's ASN.1 file.
            ----------------------
FEATURES             Location/Qualifiers
     source          1. .208197
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-459F6"
BASE COUNT    59731 a  45226 c  44706 g  58534 t
ORIGIN


  Query Match             64.0%;  Score 22.4;  DB 9;  Length 208197;
  Best Local Similarity   81.2%;  Pred. No. 40;
  Matches   26;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 catgtctagaatggggatggggtcactctgga 32
          ||| | || |||||||||| |||  |||||||
Db  98371 CATCTGTAAAATGGGGATGTGGTGCCTCTGGA 98340


RESULT  13
MMU44942
LOCUS       MMU44942                3536 bp    DNA     linear   ROD 12-APR-1996
DEFINITION  Mus musculus recombinant quaking gene sequence.
ACCESSION   U44942
VERSION     U44942.1  GI:1181701
KEYWORDS    .
SOURCE      house mouse strain=composite of DBA/2J and tw5 (a t-haplotype).
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 3536)
  AUTHORS   Ebersole,T.A., Chen,Q., Justice,M.J. and Artzt,K.
  TITLE     The quaking gene product necessary in embryogenesis and myelination
            combines features of RNA binding and signal transduction proteins
  JOURNAL   Nature Genet. 12 (3), 260-265 (1996)
  MEDLINE   96172824
REFERENCE   2  (bases 1 to 3536)
  AUTHORS   Ebersole,T.A., Chen,Qi., Tonissen,K. and Artzt,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JAN-1996) Tom A. Ebersole, Genetics, University of
            Cambridge, Downing Street, Cambridge CB2 3EH, UK
FEATURES             Location/Qualifiers
     source          1. .3536
                     /organism="Mus musculus"
                     /strain="composite of DBA/2J and tw5 (a t-haplotype)"
                     /db_xref="taxon:10090"
                     /chromosome="17"
                     /map="between T and Plg"
     misc_feature    305. .1500
                     /note="similar to QKII cDNA sequence"
     misc_feature    675
                     /note="breakpoint of qk (viable) mutation"
     misc_feature    1822. .2152
                     /note="similar to QKIII cDNA sequence; contains putative
                     SH3 binding sites"
     misc_feature    2157. .>3536
                     /note="similar to QKI cDNA sequence"
     misc_feature    2804. .3168
                     /note="similar to QKIII region 2"
BASE COUNT      766 a   1061 c   1084 g    624 t      1 others
ORIGIN


  Query Match             62.3%;  Score 21.8;  DB 10;  Length 3536;
  Best Local Similarity   92.0%;  Pred. No. 1.3e+02;
  Matches   23;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      7 tagaatggggatggggtcactctgg 31
          || |||||||||||| ||||||||
Db     31 TAAAATGGGGATGGGGGCACTCTGG 55


RESULT  14
AL137851
LOCUS       AL137851              139182 bp    DNA     linear   PRI 30-SEP-2000
DEFINITION  Human DNA sequence from clone RP11-6H18 on chromosome 9p23-24.3,
            complete sequence.
ACCESSION   AL137851
VERSION     AL137851.12  GI:10432507
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 139182)
  AUTHORS   Kimberley,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 29, 2000 this sequence version replaced gi:10178409.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            RP11-6H18 is from the library RPCI-11.1 constructed at the Roswell
            Park Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP11-6H18 The true left
            end of clone RP11-19A22 is at 96446 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .139182
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /map="p23-24.3"
                     /clone="RP11-6H18"
                     /clone_lib="RPCI-11.1"
     misc_feature    55393. .55437
                     /note="Single clone region. ALU region. Assembly confirmed
                     by restriction enzyme digest data."
BASE COUNT    46558 a  24352 c  24366 g  43906 t
ORIGIN
```

```
  Query Match             62.3%;  Score 21.8;  DB 9;  Length 139182;
  Best Local Similarity   78.8%;  Pred. No. 75;
  Matches   26;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       1 catgtctagaatggggatggggtcactctggac 33
           ||||  || |||||| ||||||| | |||| ||
Db   28549 CATGATTACAATGGGCATGGGGTAAATCTGAAC 28581


RESULT  15
HS527D15
LOCUS       HS527D15              166169 bp    DNA     linear   PRI 25-JAN-2002
DEFINITION  Homo sapiens chromosome 9 BAC RP11-527D15, complete sequence.
ACCESSION   AL513422 AL358114
VERSION     AL513422.2  GI:18378060
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Plumb,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-AUG-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
REFERENCE   2  (bases 1 to 166169)
  AUTHORS   Scharfe,M., Conrad,A., Hornischer,K., Loehnert,T.H., Thies,S. and
            Bloecker,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-FEB-2001) GBF, Dept. of Genome Analysis, Mascheroder
            Weg 1, D-38124 Braunschweig, Germany, E-mail: info.genome@gbf.de
COMMENT     On Jan 27, 2002 this sequence version replaced gi:12718190.
            All annotations in this database entry are developed by
            computational tools. It is therefore not explicitly noted in the
            feature lines that evidence is not experimental.
            Mapping was performed at The Sanger Centre
            (cf. http://www.sanger.ac.uk/HGP/Chr9)
            Mapping information is available via
            http://webace.sanger.ac.uk/cgi-bin/display?db=acedb9&grep=527D15
            -------------- Genome Center
            Center: GBF, Braunschweig
            Center code: GBF
            Web site: http://genome.gbf.de/
            Contact: info.genome@gbf.de
            -------------- Project Information
            Center project name:
            Center clone name: 527D15
            -------------- Summary Statistics
            Sequencing vector: ###;
            Chemistry: Dye-terminator-amersham: 82% of reads
            Chemistry: Dye-primer-amersham: 18% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 0 bases at least Q40
            Consensus quality: 0 bases at least Q30
            Consensus quality: 0 bases at least Q20
            Estimated insert size: ##; agarose-fp estimation
            Estimated insert size: 166169; sum-of-contigs estimation
            --------------.
            PROGRAMS AND PARAMETERS USED FOR ANNOTATION:
            ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
            +   Analysis and annotation were performed with the automatic    +
            +          'first-pass' annotation and submission tool           +
            +             'AnnoMitter' (Hornischer & Bloecker).              +
            +                 Programs used by 'AnnoMitter':                 +
            ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
            > GeneFinder (Green), Vers. 084
            .   Organism: human
            > GenScan (Burge & Karlin), Vers. 1.0
            .   Used matrix: vertebrate; Minimum score: 0
            > Grail (Xu et al.), Vers. 1.3
            .   Organism: human
            > Mzef (Zhang)
            .   Prior probability: 0.04; Overlapping number: 0 > Xpound (Thomas
            & Skolnick)
            .   Base score cutoff: 0.2; Minimal exon length: 3 bp > 'Repeats':
            BLASTN 2.0.14 (Altschul et al.)
            .   Database(s): * RepBase: ALU (human), released 22-DEC-1995 .
                      * RepBase: THR ((human), released 22-DEC-1995 .
                 * RepBase: L1 (primate), released 22-DEC-1995 .
             * RepBase: MIR (primate), released 22-DEC-1995 .                *
            RepBase: MER (primate), released 22-DEC-1995 .                  *
            RepBase: MIR2 (primate), released 22-DEC-1995 .                  *
            RepBase: THE (primate), released 22-DEC-1995 .   Minimum score: 60;
            Minimum identity: 70 %;
            > 'ESTs': BLASTN 2.0.14 (Altschul et al.)
            .   Database(s): * embl (EST, human), released -DEC- .
              * embl (EST, other), released -DEC- .                  * emblnew
            (EST), Vers. 67+ (01-JAN-1970) .   Using sequence with masked
            repeats
            .   Minimum score: 60; Minimum identity: 70 %;
            > 'Tandem Repeats': GDE 2.2 option 'tandem'
            .   Minimum length 2 bp; Maximum length 20 bp; Score threshold 20 .
              Treat N's as mismatches? YES; Allow uniform consensi? NO >
            'Inverted Repeats': GDE 2.2 option 'inverted'
            > 'Micro Satellites': GDE 2.2 option 'sputnik' (Abajian) > 'CpG
            Islands': GDE 2.2 option 'cpg'
            .   CpG island region size 100 bp;
            .   Minimum GC contents 50 %; Observed/Expected 0.6 > 'STS Scan':
            e-PCR (Schuler)
            .   Margin: 50; Number of mismatches allowed: 0; Word size: 7 .
            STS database: 'dbSTS markers'
            > 'tRNA Scan': tRNAscan-SE (Lowe & Eddy), Vers. 1.11.
FEATURES             Location/Qualifiers
     source          1. .166169
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-527D15"
                     /clone_lib="RPCI-11.2"
     misc_feature    1. .166169
                     /note="assembly_fragment~clone_end:T7~vector_side:left
                     assembly_fragment~clone_end:SP6~vector_side:right"
     repeat_region   complement(1516. .1555)
                     /note="97% identity: matches 39. .78 of consensus"
                     /rpt_family="L1"
     repeat_region   1554. .1662
                     /note="97% identity: matches 396. .504 of consensus"
                     /rpt_family="L1"
     repeat_region   2526. .2581
                     /note="homology = 75.0%, counts = 8"
                     /rpt_family="aattagt repeat"
                     /rpt_type=TANDEM
     exon            2611. .2746
                     /note="GRAIL, score = 100%, comment = excellent
                     MZEF prediction, score = 0.532"
     repeat_region   2641. .2679
                     /note="97% identity: matches 1. .39 of consensus"
                     /rpt_family="THE"
     repeat_region   2677. .3344
                     /note="81% identity: matches 274. .933 of consensus"
                     /rpt_family="THR"
     repeat_region   3405. .3443
                     /note="94% identity: matches 962. .1000 of consensus"
                     /rpt_family="THR"
     exon            complement(3433. .3468)
                     /note="XPOUND prediction, score = 0.522"
     repeat_region   3470. .3929
                     /note="85% identity: matches 1082. .1542 of consensus"
                     /rpt_family="THR"
     exon            complement(3738. .3754)
                     /note="XPOUND prediction, score = 0.227"
     repeat_region   3955. .4095
                     /note="86% identity: matches 16. .156 of consensus"
```

```
                     /rpt_family="THE"
     repeat_region   4076. .4296
                     /note="85% identity: matches 126. .338 of consensus"
                     /rpt_family="THE"
     exon            complement(4099. .4184)
                     /note="XPOUND prediction, score = 0.469"
     repeat_region   7822. .8068
                     /note="homology = 59.5%, counts = 19"
                     /rpt_family="aataaaataaaat repeat"
                     /rpt_type=TANDEM
     exon            8072. .8149
                     /note="GRAIL, score = 45%, comment = marginal"
     exon            8082. .8149
                     /note="MZEF prediction, score = 0.590"
     repeat_region   complement(8572. .9787)
                     /note="94% identity: matches 1204. .2415 of consensus"
                     /rpt_family="L1"
     exon            complement(8771. .8903)
                     /note="GRAIL, score = 52%, comment = good shadow"
     exon            complement(8815. .8903)
                     /note="XPOUND prediction, score = 0.488"
     exon            complement(9562. .9644)
                     /note="XPOUND prediction, score = 0.744"
     exon            9761. .10480
                     /note="GRAIL, score = 72%, comment = good"
     repeat_region   9782. .10894
                     /note="95% identity: matches 3666. .4779 of consensus"
                     /rpt_family="L1"
     exon            10413. .10563
                     /note="MZEF prediction, score = 0.549"
     exon            10546. .10766
                     /note="GENSCAN prediction, score = 2.06"
     exon            complement(11231. .11276)
                     /note="GRAIL, score = 77%, comment = excellent shadow"
     exon            complement(11613. .11739)
                     /note="MZEF prediction, score = 0.535"
     repeat_region   11631. .11678
                     /note="homology = 79.2%, counts = 24"
                     /rpt_family="ct repeat"
                     /rpt_type=TANDEM
     satellite       11637. .11659
                     /note="CT repeat"
     exon            complement(11897. .11903)
                     /note="XPOUND prediction, score = 0.214"
     exon            complement(11973. .12007)
                     /note="XPOUND prediction, score = 0.231"
     exon            complement(13095. .13170)
                     /note="MZEF prediction, score = 0.797"
     exon            complement(14021. .14114)
                     /note="GRAIL, score = 60%, comment = good"
     exon            complement(14684. .14743)
                     /note="MZEF prediction, score = 0.835"
     exon            16111. .16146
                     /note="MZEF prediction, score = 0.969"
     exon            16120. .16146
                     /note="GRAIL, score = 67%, comment = good shadow"
     exon            complement(16300. .16429)
                     /note="GRAIL, score = 84%, comment = excellent"
     repeat_region   17109. .17142
                     /note="homology = 94.1%, counts = 17"
                     /rpt_family="tg repeat"
                     /rpt_type=TANDEM
     satellite       17111. .17142
                     /note="TG repeat"
     satellite       17684. .17696
                     /note="TA repeat"
     repeat_region   19986. .20045

  Query Match           62.3%;  Score 21.8;  DB 9;  Length 166169;
  Best Local Similarity  78.8%;  Pred. No. 73;
  Matches   26;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       1 catgtctagaatggggatggggtcactctggac 33
                   |||| || |||||| ||||||| | |||| ||
Db   93949 CATGATTACAATGGGCATGGGGTAAATCTGAAC 93981


Search completed: June 20, 2002, 14:36:04
Job time: 19627 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2002, 14:39:39 ; Search time 668.44 Seconds
(without alignments)
89.899 Million cell updates/sec

Title: US-09-927-267-13
Perfect score: 35
Sequence: 1 catgtctagaatggggatggggtcactctggacct 35

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	25.4	72.6	1835	24	ABA04602	MOL10a coding sequ
c	2	25.4	72.6	2551	24	ABA04603	MOL10b coding sequ
	3	21.4	61.1	510	22	ABA11498	Human nervous syst
	4	21.4	61.1	523	22	ABA15447	Human nervous syst
	5	21.4	61.1	615	23	AAS69677	DNA encoding novel
	6	20.4	58.3	995	21	AAX89465	LeuD subunit of 3-
	7	20.4	58.3	1033	21	AAX89462	Corn leuD subunit
c	8	20.2	57.7	1950	23	ABL08841	Drosophila melanog
c	9	20.2	57.7	4043	23	ABL08840	Drosophila melanog
c	10	20.2	57.7	7727	22	AAL04997	Human reproductive
c	11	20.2	57.7	15339	23	ABL08078	Drosophila melanog
	12	20	57.1	41765	22	AAK76675	Human immune/haema
	13	20	57.1	41772	22	AAK76676	Human immune/haema
	14	19.8	56.6	112190	22	AAH44801	Human GPCR protein
c	15	19.6	56.0	2450	22	AAS01214	DNA encoding human
c	16	19.6	56.0	12222	22	AAF54045	Human alpha-1-anti
c	17	19.4	55.4	519	22	AAK76854	Human immune/haema
c	18	19.4	55.4	647	22	AAK60067	Human immune/haema
c	19	19.4	55.4	2948	23	ABL23532	Drosophila melanog
	20	19.2	54.9	339	21	AAF09408	Fusarium venenatum
	21	19.2	54.9	717	23	AAS87896	DNA encoding novel
	22	19.2	54.9	999	22	AAI89451	Human polynucleoti
c	23	19.2	54.9	1091	22	AAS27281	cDNA encoding nove
	24	19.2	54.9	2042	22	AAK94858	Human full-length
	25	19.2	54.9	2683	21	AAC77993	Human cancer assoc
	26	19.2	54.9	4094	21	AAZ51205	Human hypoxia regu
	27	19.2	54.9	4248	21	AAA28450	Human Seladin-1 cD
	28	19.2	54.9	4742	22	AAS42034	Genomic sequence #
	29	19.2	54.9	4742	22	AAS42035	Genomic sequence #
	30	19.2	54.9	16235	22	AAK86192	Human immune/haema
c	31	19.2	54.9	38186	20	AAZ32028	Human METH1 relate
c	32	19.2	54.9	38186	22	AAC90085	AC004449 cDNA clon
c	33	19	54.3	300	21	AAA01571	Human colon cancer
c	34	19	54.3	353	23	AAS69053	DNA encoding novel
	35	19	54.3	370	22	AAS25419	Human ovarian PCR-
c	36	19	54.3	485	16	AAT19888	Human gene signatu
c	37	19	54.3	539	22	AAK54826	Human haematologic
c	38	19	54.3	564	22	AAK54884	Human haematologic
c	39	19	54.3	576	16	AAQ83884	Hepatitis C virus
c	40	19	54.3	576	17	AAT16598	Hepatitis C virus
c	41	19	54.3	635	23	AAS58727	cDNA #1403 encodin
	42	19	54.3	732	22	ABA69539	Human foetal liver
	43	19	54.3	732	22	ABA36460	Probe #14926 for g
c	44	19	54.3	826	22	AAL01444	Human reproductive
c	45	19	54.3	879	23	AAS69057	DNA encoding novel

ALIGNMENTS

RESULT 1
ABA04602/c
ID ABA04602 standard; cDNA; 1835 BP.
XX
AC ABA04602;
XX
DT 21-FEB-2002 (first entry)
XX
DE MOL10a coding sequence.
XX
KW MOL; G-coupled protein-receptor; cardiomyopathy; atherosclerosis;
KW cell signal processing; metabolic disorder; diabetes; cancer;
KW neurodegenerative disorder; immune disorder; cardiac disorder;
KW lung disease; autoimmune disease; developmental disorder; antidiabetic;
KW Cytostatic; Neuroprotective; Antiatherosclerotic; Immunosuppressive;
KW Gene therapy; Vaccine; antiinflammatory; MOL10a; ss.
XX
OS Unidentified.
XX
FH Key Location/Qualifiers
FT CDS 54..1790
FT /*tag= a
FT /product= "MOL 10a"
XX
PN WO200181578-A2.
XX
PD 01-NOV-2001.
XX
PF 26-APR-2001; 2001WO-US13578.
XX
PR 26-APR-2000; 2000US-200158P.
PR 28-APR-2000; 2000US-200613P.

PR 28-APR-2000; 2000US-200780P.
PR 01-MAY-2000; 2000US-201006P.
PR 01-MAY-2000; 2000US-201007P.
PR 01-MAY-2000; 2000US-201236P.
PR 01-MAY-2000; 2000US-201238P.
PR 02-MAY-2000; 2000US-201186P.
PR 03-MAY-2000; 2000US-201474P.
PR 03-MAY-2000; 2000US-201508P.
PR 25-JUL-2000; 2000US-220591P.
PR 15-SEP-2000; 2000US-232678P.
PR 22-JAN-2001; 2001US-263217P.
PR 30-JAN-2001; 2001US-265160P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Vernet CAM, Fernandes ER, Gerlach V, Shimkets RA, Malyankar UM;
PI Boldog FL, Zerhusen BD, Spytek KA, Majumder K, Tchernev VT;
PI Padigaru M, Patturajan M, Burgess CE, Gangolli EA, Smithson G;
PI Rastelli L, Macdougall JR, Taupier RJ, Grosse WM, Szekeres ES;
PI Alsobrook JP;
XX
DR WPI; 2002-049278/06.
DR P-PSDB; AAM47672.
XX
PT Novel G-protein coupled receptor-related polypeptides and
PT polynucleotides for diagnosing, preventing and treating cardiomyopathy,
PT atherosclerosis, disorders related to cell signal processing and for
PT identifying modulators -
XX
PS Claim 8; Page 76; 227pp; English.
XX
CC The present invention relates to novel G-coupled protein-receptor related
CC proteins and coding sequences (MOLX, where X is a number from 1 to 10,
CC ABA04589-ABA04603 and AAM47659-AAM47673). MOLX proteins and coding
CC sequences are useful for treating or preventing a MOLX-associated
CC disorder, such as cardiomyopathy, atherosclerosis, disorders related to
CC cell signal processing and metabolic pathway modulation, diabetes and
CC cancer. Additionally, MOLX proteins and coding sequences are useful for
CC preventing and treating a variety of disorders including metabolic
CC disorders, nutritional oedema, chronic and hereditary pancreatitis,
CC obesity, infectious disease, anorexia, neurodegenerative disorders,
CC Alzheimer's disease, Parkinson's disease, stroke, immune disorders,
CC haematopoietic disorders and various dyslipidaemias, metabolic syndrome X
CC and wasting disorders associated with chronic diseases and cancers,
CC cardiac disorders, hypertension, hypercalcaemia, cirrhosis, angiogenesis
CC and wound healing, trauma, glomerulonephritis, hyper and hypothyroidism,
CC multiple sclerosis, lung diseases including asthma, Crohn's disease,
CC scleroderma, autoimmune diseases, developmental disorders and neural tube
CC defects. The present sequence is the coding sequence for MOL10a.
CC MOL10a protein is a cyclic nucleotide-gated olfactory channel-like
CC protein.
XX
SQ Sequence 1835 BP; 420 A; 521 C; 521 G; 373 T; 0 other;

Query Match 72.6%; Score 25.4; DB 24; Length 1835;
Best Local Similarity 96.3%; Pred. No. 0.44;
Matches 26; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 9 gaatggggatggggtcactctggacct 35
| ||||||||||||||||||||||||||
Db 1804 GGATGGGGATGGGGTCACTCTGGACCT 1778

RESULT 2
ABA04603/c
ID ABA04603 standard; cDNA; 2551 BP.
XX
AC ABA04603;
XX
DT 21-FEB-2002 (first entry)
XX
DE MOL10b coding sequence.
XX
KW MOL; G-coupled protein-receptor; cardiomyopathy; atherosclerosis;
KW cell signal processing; metabolic disorder; diabetes; cancer; human;
KW neurodegenerative disorder; immune disorder; cardiac disorder;
KW lung disease; autoimmune disease; developmental disorder; antidiabetic;
KW Cytostatic; Neuroprotective; Antiatherosclerotic; Immunosuppressive;
KW Gene therapy; Vaccine; antiinflammatory; MOL10b; chromosome 11; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 779..2506
FT /*tag= a
FT /product= "MOL 10b"
XX
PN WO200181578-A2.
XX
PD 01-NOV-2001.
XX
PF 26-APR-2001; 2001WO-US13578.
XX
PR 26-APR-2000; 2000US-200158P.
PR 28-APR-2000; 2000US-200613P.
PR 28-APR-2000; 2000US-200780P.
PR 01-MAY-2000; 2000US-201006P.
PR 01-MAY-2000; 2000US-201007P.
PR 01-MAY-2000; 2000US-201236P.
PR 01-MAY-2000; 2000US-201238P.
PR 02-MAY-2000; 2000US-201186P.
PR 03-MAY-2000; 2000US-201474P.
PR 03-MAY-2000; 2000US-201508P.
PR 25-JUL-2000; 2000US-220591P.
PR 15-SEP-2000; 2000US-232678P.
PR 22-JAN-2001; 2001US-263217P.
PR 30-JAN-2001; 2001US-265160P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Vernet CAM, Fernandes ER, Gerlach V, Shimkets RA, Malyankar UM;
PI Boldog FL, Zerhusen BD, Spytek KA, Majumder K, Tchernev VT;
PI Padigaru M, Patturajan M, Burgess CE, Gangolli EA, Smithson G;
PI Rastelli L, Macdougall JR, Taupier RJ, Grosse WM, Szekeres ES;
PI Alsobrook JP;
XX
DR WPI; 2002-049278/06.
DR P-PSDB; AAM47673.
XX
PT Novel G-protein coupled receptor-related polypeptides and
PT polynucleotides for diagnosing, preventing and treating cardiomyopathy,
PT atherosclerosis, disorders related to cell signal processing and for
PT identifying modulators -
XX
PS Claim 8; Page 78; 227pp; English.
XX
CC The present invention relates to novel G-coupled protein-receptor related
CC proteins and coding sequences (MOLX, where X is a number from 1 to 10,
CC ABA04589-ABA04603 and AAM47659-AAM47673). MOLX proteins and coding
CC sequences are useful for treating or preventing a MOLX-associated
CC disorder, such as cardiomyopathy, atherosclerosis, disorders related to
CC cell signal processing and metabolic pathway modulation, diabetes and
CC cancer. Additionally, MOLX proteins and coding sequences are useful for
CC preventing and treating a variety of disorders including metabolic
CC disorders, nutritional oedema, chronic and hereditary pancreatitis,
CC obesity, infectious disease, anorexia, neurodegenerative disorders,
CC Alzheimer's disease, Parkinson's disease, stroke, immune disorders,
CC haematopoietic disorders and various dyslipidaemias, metabolic syndrome X
CC and wasting disorders associated with chronic diseases and cancers,
CC cardiac disorders, hypertension, hypercalcaemia, cirrhosis, angiogenesis
CC and wound healing, trauma, glomerulonephritis, hyper and hypothyroidism,
CC multiple sclerosis, lung diseases including asthma, Crohn's disease,
CC scleroderma, autoimmune diseases, developmental disorders and neural tube
CC defects. The present sequence is the coding sequence for human MOL10b,
CC which maps to chromosome 11. MOL10b protein is a cyclic nucleotide-gated

```
CC   olfactory channel-like protein.
XX
SQ   Sequence 2551 BP; 646 A; 668 C; 707 G; 530 T; 0 other;


  Query Match            72.6%;  Score 25.4;  DB 24;  Length 2551;
  Best Local Similarity  96.3%;  Pred. No. 0.46;
  Matches   26;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       9 gaatggggatggggtcactctggacct 35
           | ||||||||||||||||||||||||||
Db    2520 GGATGGGGATGGGGTCACTCTGGACCT 2494


RESULT   3
ABA11498
ID   ABA11498 standard; cDNA; 510 BP.
XX
AC   ABA11498;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Human nervous system related polynucleotide SEQ ID NO 505.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine; ss.
XX
OS   Homo sapiens.
XX
PN   WO200159063-A2.
XX
PD   16-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01334.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   20-OCT-2000; 2000US-0242221.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
```

PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-541565/60.
DR P-PSDB; ABB15172.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
PS Claim 1; SEQ ID NO 505; 1701pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 510 BP; 135 A; 101 C; 116 G; 157 T; 1 other;

Query Match 61.1%; Score 21.4; DB 22; Length 510;
Best Local Similarity 80.6%; Pred. No. 16;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 catgtctagaatggggatggggtcactctgg 31
|||| || || |||||||||||| ||||||
Db 84 catggctgaaagagggatggggtctctctgg 114

RESULT 4
ABA15447
ID ABA15447 standard; DNA; 523 BP.
XX
AC ABA15447;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 7778.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01334.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.

PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-541565/60.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
PS Disclosure; SEQ ID NO 7778; 1701pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 523 BP; 136 A; 103 C; 119 G; 165 T; 0 other;

Query Match 61.1%; Score 21.4; DB 22; Length 523;
Best Local Similarity 80.6%; Pred. No. 16;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 catgtctagaatggggatggggtcactctgg 31
|||| || || |||||||||| ||||||
Db 84 catggctgaaagagggatggggtctctctgg 114

RESULT 5
AAS69677
ID AAS69677 standard; cDNA; 615 BP.
XX
AC AAS69677;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #5481.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG05490.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 5481; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 615 BP; 187 A; 158 C; 147 G; 123 T; 0 other;

Query Match 61.1%; Score 21.4; DB 23; Length 615;
Best Local Similarity 80.6%; Pred. No. 17;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 5 tctagaatggggatggggtcactctggacct 35
||| |||||||||| |||||||| | || |
Db 101 tctccaatggggatgtggtcactcagtactt 131

RESULT 6
AAX89465
ID AAX89465 standard; cDNA; 995 BP.
XX
AC AAX89465;
XX
DT 15-FEB-2000 (first entry)
XX
DE LeuD subunit of 3-isopropylmalate dehydratase nucleotide sequence.
XX
KW Corn; soybean; wheat; rice; dihydroxyacid dehydratase; leuC; leuD;
KW branched chain amino acid transferase; biosynthetic enzyme; antibody;
KW 3-isopropylmalate dehydratase; ss.
XX
OS Triticum aestivum.
XX
FH Key Location/Qualifiers
FT CDS 1..735
FT /*tag= a
FT /product= LeuD
XX
PN WO9921880-A2.
XX
PD 06-MAY-1999.
XX
PF 20-OCT-1998; 98WO-US22081.
XX
PR 28-OCT-1997; 97US-0063423.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Falco SC, Cahoon RE, Hitz WD, Kinney AJ, Rafalski JA;
XX
DR WPI; 2000-022904/02.
DR P-PSDB; AAY28442.
XX
PT Nucleic acid fragments encoding branched chain amino acid biosynthetic
PT enzymes -
XX
PS Claim 17; Page 98; 102pp; English.
XX
CC AAX89462-X89465 are fragments of corn, soybean, wheat and rice leuD
CC subunit of 3-isopropylmalate dehydratase nucleotide sequences. Sequences
CC AAX89442-X89465 are nucleic acid fragments that encode all or a
CC substantial portion of dihydroxyacid dehydratase, a branched chain amino
CC acid transferase, a leuC or a leuD subunit of 3-isopropylmalate
CC dehydratase from wheat, corn, soybean or rice. These enzymes are involved
CC in biosynthesis and utilization of branched-chain amino acids. The
CC nucleic acid sequences can be used to alter the level of expression of a
CC branched chain amino acid biosynthetic enzyme in a host cell. They can
CC also be used to obtain a nucleic acid fragment encoding a branched chain
CC amino acid biosynthetic enzyme, and to identify inhibitors of a branched
CC chain amino acid biosynthetic enzyme. The encoded proteins may be used
CC to prepare antibodies for detecting the proteins in situ in cells, or in
CC vitro in cell extracts.
XX
SQ Sequence 995 BP; 194 A; 296 C; 277 G; 228 T; 0 other;

Query Match 58.3%; Score 20.4; DB 21; Length 995;
Best Local Similarity 80.0%; Pred. No. 47;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 6 ctagaatggggatggggtcactctggacct 35
| ||| ||||||| |||||| |||||||
Db 561 caagacaggggatgtggtcaccgtggacct 590

RESULT 7
AAX89462
ID AAX89462 standard; cDNA; 1033 BP.
XX
AC AAX89462;
XX
DT 15-FEB-2000 (first entry)
XX
DE Corn leuD subunit of 3-isopropylmalate dehydratase nucleotide sequence.
XX
KW Corn; soybean; wheat; rice; dihydroxyacid dehydratase; leuC; leuD;
KW branched chain amino acid transferase; biosynthetic enzyme; antibody;
KW 3-isopropylmalate dehydratase; ss.
XX
OS Zea mays.
XX
FH Key Location/Qualifiers

```
FT   CDS             4..753
FT                   /*tag=  a
FT                   /product= LeuD
FT                   /note= "Subunit of 3-isopropylmalate dehydratase"
XX
PN   WO9921880-A2.
XX
PD   06-MAY-1999.
XX
PF   20-OCT-1998;   98WO-US22081.
XX
PR   28-OCT-1997;   97US-0063423.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Falco SC,  Cahoon RE,  Hitz WD,  Kinney AJ,  Rafalski JA;
XX
DR   WPI; 2000-022904/02.
DR   P-PSDB; AAY28439.
XX
PT   Nucleic acid fragments encoding branched chain amino acid biosynthetic
PT   enzymes  -
XX
PS   Claim 17; Page 94-95; 102pp; English.
XX
CC   AAX89462-X89465 are fragments of corn, soybean, wheat and rice leuD
CC   subunit of 3-isopropylmalate dehydratase nucleotide sequences. Sequences
CC   AAX89442-X89465 are nucleic acid fragments that encode all or a
CC   substantial portion of dihydroxyacid dehydratase, a branched chain amino
CC   acid transferase, a leuC or a leuD subunit of 3-isopropylmalate
CC   dehydratase from wheat, corn, soybean or rice. These enzymes are involved
CC   in biosynthesis and utilization of branched-chain amino acids. The
CC   nucleic acid sequences can be used to alter the level of expression of a
CC   branched chain amino acid biosynthetic enzyme in a host cell. They can
CC   also be used to obtain a nucleic acid fragment encoding a branched chain
CC   amino acid biosynthetic enzyme, and to identify inhibitors of a branched
CC   chain amino acid biosynthetic enzyme. The encoded proteins may be used
CC   to prepare antibodies for detecting the proteins in situ in cells, or in
CC   vitro in cell extracts.
XX
SQ   Sequence 1033 BP; 221 A; 299 C; 282 G; 230 T; 1 other;


  Query Match             58.3%;  Score 20.4;  DB 21;  Length 1033;
  Best Local Similarity   80.0%;  Pred. No. 48;
  Matches   24;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      6 ctagaatggggatggggtcactctggacct 35
          | |||  ||||||| ||||||  |||||||
Db    579 caagacaggggatgtggtcaccgtggacct 608


RESULT   8
ABL08841/c
ID   ABL08841 standard; cDNA; 1950 BP.
XX
AC   ABL08841;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 21005.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB64738.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 21005; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1950 BP; 455 A; 524 C; 553 G; 418 T; 0 other;


  Query Match             57.7%;  Score 20.2;  DB 23;  Length 1950;
  Best Local Similarity   75.8%;  Pred. No. 64;
  Matches   25;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      1 catgtctagaatggggatggggtcactctggac 33
          |||||| |||| |  |||||||| ||  || ||
Db   1467 CATGTCCAGAAAGTAGATGGGGTAACAGTGAAC 1435


RESULT   9
ABL08840/c
ID   ABL08840 standard; cDNA; 4043 BP.
XX
AC   ABL08840;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 21002.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB64737.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
```

```
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 21002; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 4043 BP; 1222 A; 870 C; 909 G; 1042 T; 0 other;


  Query Match             57.7%;  Score 20.2;  DB 23;  Length 4043;
  Best Local Similarity   75.8%;  Pred. No. 72;
  Matches   25;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       1 catgtctagaatggggatggggtcactctggac 33
           |||||| |||| |  |||||||| ||  || ||
Db    2560 CATGTCCAGAAAGTAGATGGGGTAACAGTGAAC 2528


RESULT  10
AAL04997/c
ID   AAL04997 standard; DNA; 7727 BP.
XX
AC   AAL04997;
XX
DT   21-NOV-2001  (first entry)
XX
DE   Human reproductive system related antigen DNA SEQ ID NO: 7685.
XX
KW   Human; reproductive system related antigen; reproductive system disorder;
KW   cancer; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155320-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01339.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
```

PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465570/50.
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen
PT is used in preventing, treating or ameliorating a medical condition -
XX
PS Disclosure; SEQ ID NO 7685; 1297pp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.
XX
SQ Sequence 7727 BP; 2170 A; 1681 C; 1765 G; 2111 T; 0 other;

Query Match 57.7%; Score 20.2; DB 22; Length 7727;
Best Local Similarity 75.8%; Pred. No. 80;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 catgtctagaatggggatggggtcactctggac 33
|||| |||| |||||| |||| ||||| | |
Db 3601 CATGGCTAGCTTGGGGAGGGGGATACTCTTGGC 3569

RESULT 11
ABL08078/c
ID ABL08078 standard; cDNA; 15339 BP.
XX
AC ABL08078;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 18716.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB63975.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 18716; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 15339 BP; 4523 A; 3318 C; 3383 G; 4115 T; 0 other;

Query Match 57.7%; Score 20.2; DB 23; Length 15339;
Best Local Similarity 75.8%; Pred. No. 90;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 catgtctagaatggggatggggtcactctggac 33
|||||| |||| | |||||||| || || ||
Db 9749 CATGTCCAGAAAGTAGATGGGGTAACAGTGAAC 9717

RESULT 12
AAK76675
ID AAK76675 standard; DNA; 41765 BP.
XX
AC AAK76675;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:31487.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.

XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964:
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX

```
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Disclosure; SEQ ID NO 31487; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 41765 BP; 10094 A; 9887 C; 10072 G; 11712 T; 0 other;


  Query Match             57.1%;  Score 20;  DB 22;  Length 41765;
  Best Local Similarity   82.1%;  Pred. No. 1.3e+02;
  Matches   23;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       6 ctagaatggggatggggtcactctggac 33
           |||| ||||  ||||||||| ||||| |
Db   28751 ctagcatggaaatggggtcaatctgggc 28778


RESULT  13
AAK76676
ID   AAK76676 standard; DNA; 41772 BP.
XX
AC   AAK76676;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:31488.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
```

PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 31488; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 41772 BP; 10094 A; 9887 C; 10074 G; 11717 T; 0 other;

Query Match 57.1%; Score 20; DB 22; Length 41772;
Best Local Similarity 82.1%; Pred. No. 1.3e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 6 ctagaatggggatggggtcactctggac 33
|||| |||| ||||||||| ||||| |
Db 28759 ctagcatggaaatggggtcaatctgggc 28786

RESULT 14
AAH44801
ID AAH44801 standard; DNA; 112190 BP.
XX
AC AAH44801;
XX
DT 19-DEC-2001 (first entry)
XX
DE Human GPCR protein KAT06734L DNA containing exons 3 to 9 SEQ ID NO:16.
XX
KW Human; G protein-coupled receptor; GPCR; KAT06734L; drug development;
KW ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT exon 31560..31663
FT /*tag= a
FT /number= 3
FT exon 64868..64961
FT /*tag= b
FT /number= 4
FT exon 80499..80700
FT /*tag= c
FT /number= 5
FT exon 87482..87558
FT /*tag= d
FT /number= 6
FT exon 97994..98080
FT /*tag= e
FT /number= 7
FT exon 101581..101761
FT /*tag= f
FT /number= 8
FT exon 102663..103161
FT /*tag= g
FT /number= 9
XX
PN JP2001245666-A.
XX
PD 11-SEP-2001.
XX
PF 06-MAR-2000; 2000JP-0060548.
XX
PR 06-MAR-2000; 2000JP-0060548.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
DR WPI; 2001-610074/70.
XX
PT New G protein-coupled receptor polypeptide for use in the development
PT of new drugs -
XX
PS Example 3; Page 74-114; 126pp; Japanese.

```
XX
CC   The present invention describes a human G protein-coupled receptor
CC   (GPCR) designated KAT06734L protein. The GPCR protein can be used for
CC   the development of new drugs. The present sequence represents a
CC   KAT06734L DNA sequence containing exons 3 to 9, which is given in an
CC   example from the present invention.
XX
SQ   Sequence 112190 BP; 36965 A; 22192 C; 21827 G; 31206 T; 0 other;


  Query Match             56.6%;  Score 19.8;  DB 22;  Length 112190;
  Best Local Similarity   77.4%;  Pred. No. 1.8e+02;
  Matches   24;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       5 tctagaatggggatggggtcactctggacct 35
            ||||| |   ||| |||| |||| |||||||
Db   87750 tctaggagaaggaagggggcactttggacct 87780


RESULT  15
AAS01214/c
ID   AAS01214 standard; cDNA; 2450 BP.
XX
AC   AAS01214;
XX
DT   04-JUL-2001  (first entry)
XX
DE   DNA encoding human secreted protein, POLY5.
XX
KW   Human secreted protein; therapeutic; diagnostic; human; cancer; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             38..1540
FT                   /*tag= a
FT                   /product= "Human secreted protein POLY5"
XX
PN   WO200119856-A2.
XX
PD   22-MAR-2001.
XX
PF   13-SEP-2000; 2000WO-US25106.
XX
PR   13-SEP-1999;   99US-0153629.
PR   16-SEP-1999;   99US-0154520.
PR   20-SEP-1999;   99US-0154762.
PR   13-OCT-1999;   99US-0159231.
PR   12-SEP-2000; 2000US-0659634.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Fernandes E,  Herrmann JL,  Liu X,  Yang M,  Boldog FL;
XX
DR   WPI; 2001-244781/25.
DR   P-PSDB; AAU00393.
XX
PT   New POLYX polypeptide useful for treating or preventing a POLYX
PT   associated disorder, e.g. cancer -
XX
PS   Claim 9; Page 16-18; 152pp; English.
XX
CC   The sequence represents the coding sequence of human secreted protein,
CC   POLY5. POLYX nucleic acids, polypeptides and antibodies to POLYX can be
CC   used for treating or preventing a POLYX associated disorder in a subject,
CC   preferably a human. These can be used in the manufacture of a medicament
CC   for treating a syndrome associated with a human disease selected from a
CC   POLYX-associated disorder, where the therapeutic is a POLYX polypeptide,
CC   a POLYX nucleotide or a POLYX antibody. They may also be used to screen
CC   for a modulator of activity, or latency, or predisposition to a POLYX
CC   associated disorder, e.g. cancer.
XX
SQ   Sequence 2450 BP; 786 A; 463 C; 488 G; 713 T; 0 other;


  Query Match             56.0%;  Score 19.6;  DB 22;  Length 2450;
  Best Local Similarity   73.5%;  Pred. No. 1.2e+02;
  Matches   25;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       1 catgtctagaatggggatggggtcactctggacc 34
           | | |||||| ||||||| |  ||| |||  |||
Db    1078 CTTCTCTAGATTGGGGATCGATTCATTCTTCACC 1045


Search completed: June 20, 2002, 14:39:44
Job time: 13582 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2002, 14:28:42 ; Search time 155.16 Seconds
(without alignments)
55.408 Million cell updates/sec

Title: US-09-927-267-13
Perfect score: 35
Sequence: 1 catgtctagaatggggatggggtcactctggacct 35

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	19	54.3	576	1	US-08-086-428B-40	Sequence 40, Appl
c	2	19	54.3	576	2	US-08-468-570-40	Sequence 40, Appl
c	3	19	54.3	576	2	US-08-290-665A-40	Sequence 40, Appl
c	4	19	54.3	576	5	PCT-US95-10398-40	Sequence 40, Appl
c	5	19	54.3	1584	1	US-08-748-485-2	Sequence 2, Appli
c	6	19	54.3	2744	3	US-09-071-101-1	Sequence 1, Appli
c	7	19	54.3	2744	3	US-09-369-618-1	Sequence 1, Appli
c	8	19	54.3	2744	3	US-09-369-617-1	Sequence 1, Appli
c	9	18.8	53.7	688	5	PCT-US94-04361-26	Sequence 26, Appl
c	10	18.8	53.7	3560	1	US-08-121-713D-59	Sequence 59, Appl
c	11	18.8	53.7	3560	1	US-08-835-268-59	Sequence 59, Appl
c	12	18.8	53.7	3560	2	US-09-060-692-59	Sequence 59, Appl
c	13	18.8	53.7	3560	3	US-08-833-391-59	Sequence 59, Appl
c	14	18.8	53.7	3560	4	US-09-060-610-59	Sequence 59, Appl
c	15	18.8	53.7	3560	5	PCT-US94-10151A-59	Sequence 59, Appl
c	16	18.6	53.1	570	4	US-09-091-725-24	Sequence 24, Appl
c	17	18.6	53.1	1946	2	US-08-755-584-1	Sequence 1, Appli
c	18	18.6	53.1	1946	3	US-09-192-611-1	Sequence 1, Appli
c	19	18.2	52.0	46	6	5196308-1	Patent No. 5196308
c	20	18	51.4	1305	2	US-08-896-410-37	Sequence 37, Appl
c	21	18	51.4	1974	1	US-08-625-322-1	Sequence 1, Appli
	22	18	51.4	2557	4	US-08-464-954A-1	Sequence 1, Appli
c	23	17.8	50.9	147	2	US-08-676-279-53	Sequence 53, Appl
c	24	17.8	50.9	196	2	US-08-676-279-35	Sequence 35, Appl
c	25	17.8	50.9	687	5	PCT-US94-04361-25	Sequence 25, Appl
c	26	17.8	50.9	2007	4	US-08-637-823B-29	Sequence 29, Appl
c	27	17.8	50.9	2008	4	US-08-637-823B-3	Sequence 3, Appli
	28	17.6	50.3	150	1	US-07-969-931-14	Sequence 14, Appl
	29	17.6	50.3	150	1	US-07-855-417A-14	Sequence 14, Appl
	30	17.6	50.3	627	3	US-08-464-052-3	Sequence 3, Appli
	31	17.6	50.3	627	4	US-08-461-002-3	Sequence 3, Appli
	32	17.6	50.3	627	4	US-08-689-411-3	Sequence 3, Appli
c	33	17.6	50.3	877	4	US-08-998-416-501	Sequence 501, App
c	34	17.6	50.3	1059	2	US-08-726-306A-16	Sequence 16, Appl
c	35	17.6	50.3	1107	1	US-08-159-969-1	Sequence 1, Appli
c	36	17.6	50.3	1173	2	US-08-244-951A-9	Sequence 9, Appli
c	37	17.6	50.3	1173	2	US-08-389-011-24	Sequence 24, Appl
c	38	17.6	50.3	1173	3	US-08-403-917A-24	Sequence 24, Appl
c	39	17.6	50.3	1173	4	US-09-348-952A-24	Sequence 24, Appl
c	40	17.6	50.3	1445	4	US-09-372-422A-1	Sequence 1, Appli
c	41	17.6	50.3	1445	4	US-09-372-448A-1	Sequence 1, Appli
	42	17.6	50.3	1535	3	US-08-464-052-1	Sequence 1, Appli
	43	17.6	50.3	1535	4	US-08-461-002-1	Sequence 1, Appli
	44	17.6	50.3	1535	4	US-08-689-411-1	Sequence 1, Appli
	45	17.6	50.3	1535	5	PCT-US94-09863-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT   1
US-08-086-428B-40/c
; Sequence 40, Application US/08086428B
; Patent No. 5514539
;  GENERAL INFORMATION:
;    APPLICANT:  BUKH, J., MILLER, R.H. AND
;    APPLICANT:  PURCELL, R.H.
;    TITLE OF INVENTION:  NUCLEOTIDE AND DEDUCED
;    TITLE OF INVENTION:  AMINO ACID SEQUENCES OF THE ENVELOPE 1 GENE
;    TITLE OF INVENTION:  OF 51 ISOLATES OF HEPATITIS C AND THE USE
;    TITLE OF INVENTION:  OF REAGENTS DERIVED FROM THESE SEQUENCES IN
;    TITLE OF INVENTION:  DIAGNOSTIC METHODS AND VACCINES
;    NUMBER OF SEQUENCES:  159
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/086,428B
;      FILING DATE:  29-JUN-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  RICHARD W. BORK
;      REGISTRATION NUMBER:  36,459
;      REFERENCE/DOCKET NUMBER:  2026-4070
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;      TELEX:  421792
;  INFORMATION FOR SEQ ID NO:  40:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  576 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    ORIGINAL SOURCE:
;      ORGANISM:  homosapiens
;      INDIVIDUAL ISOLATE:  Z4
US-08-086-428B-40


  Query Match            54.3%;  Score 19;  DB 1;  Length 576;
```

```
  Best Local Similarity   71.4%;  Pred. No. 19;
  Matches   25;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      1 catgtctagaatggggatggggtcactctggacct 35
          |||||| |||| ||   ||||| | |||  || ||
Db    468 CATGTCGAGAAAGGCTGTGGGGACCCTCATGATCT 434


RESULT   2
US-08-468-570-40/c
; Sequence 40, Application US/08468570
; Patent No. 5871962
;  GENERAL INFORMATION:
;    APPLICANT:  BUKH, J., MILLER, R.H. AND
;    APPLICANT:  PURCELL, R.H.
;    TITLE OF INVENTION:  NUCLEOTIDE AND DEDUCED
;    TITLE OF INVENTION:  AMINO ACID SEQUENCES OF THE ENVELOPE 1 GENE
;    TITLE OF INVENTION:  OF 51 ISOLATES OF HEPATITIS C AND THE USE
;    TITLE OF INVENTION:  OF REAGENTS DERIVED FROM THESE SEQUENCES IN
;    TITLE OF INVENTION:  DIAGNOSTIC METHODS AND VACCINES
;    NUMBER OF SEQUENCES:  159
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/468,570
;      FILING DATE:  6-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/086,428
;      FILING DATE:  29-JUN-1993
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  RICHARD W. BORK
;      REGISTRATION NUMBER:  36,459
;      REFERENCE/DOCKET NUMBER:  2026-4070US1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;      TELEX:  421792
;  INFORMATION FOR SEQ ID NO:  40:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  576 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    ORIGINAL SOURCE:
;      ORGANISM:  homosapiens
;      INDIVIDUAL ISOLATE:  Z4
US-08-468-570-40


  Query Match             54.3%;  Score 19;  DB 2;  Length 576;
  Best Local Similarity   71.4%;  Pred. No. 19;
  Matches   25;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      1 catgtctagaatggggatggggtcactctggacct 35
          |||||| |||| ||   ||||| | |||  || ||
Db    468 CATGTCGAGAAAGGCTGTGGGGACCCTCATGATCT 434


RESULT   3
US-08-290-665A-40/c
; Sequence 40, Application US/08290665A
; Patent No. 5882852
;  GENERAL INFORMATION:
;    APPLICANT:  BUKH, J., MILLER, R.H. AND
;    APPLICANT:  PURCELL, R.H.
;    TITLE OF INVENTION:  NUCLEOTIDE AND DEDUCED
;    TITLE OF INVENTION:  AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
;    TITLE OF INVENTION:  CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
;    TITLE OF INVENTION:  AND THE USE OF REAGENTS DERIVED FROM THESE
;    TITLE OF INVENTION:  SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
;    NUMBER OF SEQUENCES:  263
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/290,665A
;      FILING DATE:  15-AUG-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  RICHARD W. BORK
;      REGISTRATION NUMBER:  36,459
;      REFERENCE/DOCKET NUMBER:  2026-4116
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;      TELEX:  421792
;  INFORMATION FOR SEQ ID NO:  40:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  576 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    ORIGINAL SOURCE:
;      ORGANISM:  homosapiens
;      INDIVIDUAL ISOLATE:  Z4
US-08-290-665A-40


  Query Match             54.3%;  Score 19;  DB 2;  Length 576;
  Best Local Similarity   71.4%;  Pred. No. 19;
  Matches   25;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      1 catgtctagaatggggatggggtcactctggacct 35
          |||||| |||| ||   ||||| | |||  || ||
Db    468 CATGTCGAGAAAGGCTGTGGGGACCCTCATGATCT 434


RESULT   4
PCT-US95-10398-40/c
; Sequence 40, Application PC/TUS9510398
;  GENERAL INFORMATION:
;    APPLICANT:  BUKH, J., MILLER, R.H. AND
;    APPLICANT:  PURCELL, R.H.
;    TITLE OF INVENTION:  NUCLEOTIDE AND DEDUCED
;    TITLE OF INVENTION:  AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
;    TITLE OF INVENTION:  CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
;    TITLE OF INVENTION:  AND THE USE OF REAGENTS DERIVED FROM THESE
;    TITLE OF INVENTION:  SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
;    NUMBER OF SEQUENCES:  263
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
```

```
;       ZIP:  10154
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  FLOPPY DISK
;       COMPUTER:  IBM PC COMPATIBLE
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  WORDPERFECT 5.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US95/10398
;       FILING DATE:  15-AUG-1995
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/086,428
;       FILING DATE:  29 JUNE 1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/290/665
;       FILING DATE:  15 AUGUST 1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  RICHARD W. BORK
;       REGISTRATION NUMBER:  36,459
;       REFERENCE/DOCKET NUMBER:  2026-4116
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 758-4800
;       TELEFAX:  (212) 751-6849
;       TELEX:  421792
;   INFORMATION FOR SEQ ID NO:  40:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  576 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     ORIGINAL SOURCE:
;       ORGANISM:  homosapiens
;       INDIVIDUAL ISOLATE:  Z4
PCT-US95-10398-40


  Query Match             54.3%;  Score 19;  DB 5;  Length 576;
  Best Local Similarity   71.4%;  Pred. No. 19;
  Matches   25;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      1 catgtctagaatggggatggggtcactctggacct 35
          |||||| |||| ||   ||||| | |||  || ||
Db    468 CATGTCGAGAAAGGCTGTGGGGACCCTCATGATCT 434


RESULT   5
US-08-748-485-2/c
; Sequence 2, Application US/08748485
; Patent No. 5817480
;  GENERAL INFORMATION:
;     APPLICANT:  Au-Young, Janice
;     APPLICANT:  Guegler, Karl J.
;     APPLICANT:  Goli, Surya K.
;     APPLICANT:  Murry, Lynn E.
;     TITLE OF INVENTION:  NOVEL HISTAMINE H2 RECEPTOR
;     NUMBER OF SEQUENCES:  8
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  INCYTE PHARMACEUTICALS, INC.
;       STREET:  3174 Porter Drive
;       CITY:  Palo Alto
;       STATE:  CA
;       COUNTRY:  US
;       ZIP:  94304
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/748,485
;       FILING DATE:  Herewith
;       CLASSIFICATION:  530
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Billings, Lucy J.
;       REGISTRATION NUMBER:  36,749
;       REFERENCE/DOCKET NUMBER:  PF-0159 US
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  415-855-0555
;       TELEFAX:  415-845-4166
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1584 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     IMMEDIATE SOURCE:
;       LIBRARY:  Consensus
;       CLONE:  1722180
US-08-748-485-2


  Query Match             54.3%;  Score 19;  DB 1;  Length 1584;
  Best Local Similarity   71.4%;  Pred. No. 24;
  Matches   25;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      1 catgtctagaatggggatggggtcactctggacct 35
          |||| | |||| ||||| | || ||| |  || ||
Db    778 CATGACCAGAAAGGGGAAGAGGGCACACCAGATCT 744


RESULT   6
US-09-071-101-1/c
; Sequence 1, Application US/09071101
; Patent No. 6013503
;  GENERAL INFORMATION:
;     APPLICANT:  Lok, Si
;     APPLICANT:  Jaspers, Stephen R.
;     TITLE OF INVENTION:  HUMAN PROHORMONE CONVERTASE 4
;     NUMBER OF SEQUENCES:  11
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  ZymoGenetics, Inc.
;       STREET:  1201 Eastlake Avenue East
;       CITY:  Seattle
;       STATE:  WA
;       COUNTRY:  USA
;       ZIP:  98102
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/071,101
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Parker, Gary E.
;       REGISTRATION NUMBER:  31,648
;       REFERENCE/DOCKET NUMBER:  97-05
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  206-442-6673
;       TELEFAX:  206-442-6678
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2744 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  61...2325
;      OTHER INFORMATION:
US-09-071-101-1


  Query Match             54.3%;  Score 19;  DB 3;  Length 2744;
  Best Local Similarity   81.5%;  Pred. No. 26;
  Matches   22;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      8 agaatggggatggggtcactctggacc 34
          || |||||| |||||   |||||||||
Db   1460 AGGATGGGGGTGGGGCGGCTCTGGACC 1434


RESULT   7
US-09-369-618-1/c
; Sequence 1, Application US/09369618
; Patent No. 6100041
; GENERAL INFORMATION:
;  APPLICANT: Lok, Si
;  APPLICANT: Jaspers, Stephen R.
;  TITLE OF INVENTION: HUMAN PROHORMONE CONVERTASE 4
;  FILE REFERENCE: 97-05D2
;  CURRENT APPLICATION NUMBER: US/09/369,618
;  CURRENT FILING DATE: 1999-08-06
;  EARLIER APPLICATION NUMBER: US 09/071,101
;  EARLIER FILING DATE: 1998-05-01
;  EARLIER APPLICATION NUMBER: US 60/044,015
;  EARLIER FILING DATE: 1998-05-06
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 2744
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (61)...(2325)
US-09-369-618-1


  Query Match             54.3%;  Score 19;  DB 3;  Length 2744;
  Best Local Similarity   81.5%;  Pred. No. 26;
  Matches   22;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      8 agaatggggatggggtcactctggacc 34
          || |||||| |||||   |||||||||
Db   1460 AGGATGGGGGTGGGGCGGCTCTGGACC 1434


RESULT   8
US-09-369-617-1/c
; Sequence 1, Application US/09369617
; Patent No. 6127162
; GENERAL INFORMATION:
;  APPLICANT: Lok, Si
;  APPLICANT: Jaspers, Stephen R.
;  TITLE OF INVENTION: HUMAN PROHORMONE CONVERTASE 4
;  FILE REFERENCE: 97-05D1
;  CURRENT APPLICATION NUMBER: US/09/369,617
;  CURRENT FILING DATE: 1999-08-06
;  EARLIER APPLICATION NUMBER: US 09/071,101
;  EARLIER FILING DATE: 1998-05-01
;  EARLIER APPLICATION NUMBER: US 60/044,015
;  EARLIER FILING DATE: 1998-05-06
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 2744
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (61)...(2325)
US-09-369-617-1


  Query Match             54.3%;  Score 19;  DB 3;  Length 2744;
  Best Local Similarity   81.5%;  Pred. No. 26;
  Matches   22;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      8 agaatggggatggggtcactctggacc 34
          || |||||| |||||   |||||||||
Db   1460 AGGATGGGGGTGGGGCGGCTCTGGACC 1434


RESULT   9
PCT-US94-04361-26/c
; Sequence 26, Application PC/TUS9404361
;  GENERAL INFORMATION:
;    APPLICANT:  Brigham and Women's Hospital
;    APPLICANT:  75 Francis Street
;    APPLICANT:  Boston, MA 02115
;    APPLICANT:  Bunn, H. Franklin
;    APPLICANT:  Wen, Danyi
;    APPLICANT:  Showers, Mark O.
;    TITLE OF INVENTION:  Erythropoietin Muteins With Enhanced
;    TITLE OF INVENTION:  Activity
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox
;      STREET:  1100 New York Avenue, Suite 600
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/04361
;      FILING DATE:  Herewith
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/049,802
;      FILING DATE:  21-APR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cimbala, Michele A.
;      REGISTRATION NUMBER:  33,851
;      REFERENCE/DOCKET NUMBER:  0627.336PC01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  688 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  both
PCT-US94-04361-26


  Query Match             53.7%;  Score 18.8;  DB 5;  Length 688;
  Best Local Similarity   76.7%;  Pred. No. 24;
  Matches   23;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      4 gtctagaatggggatggggtcactctggac 33
          |||| ||| ||||||||  ||  ||||| |
Db    435 GTCTGGAAGGGGGATGGCTTCCTTCTGGGC 406
```

```
RESULT  10
US-08-121-713D-59/c
; Sequence 59, Application US/08121713D
; Patent No. 5639856
;  GENERAL INFORMATION:
;    APPLICANT:  Goodman, Corey S.
;    APPLICANT:  Kolodkin, Alex L.
;    APPLICANT:  Matthes, David
;    APPLICANT:  Bentley, David R.
;    APPLICANT:  O'Connor, Timothy
;    TITLE OF INVENTION:  The Semaphorin Gene Family
;    NUMBER OF SEQUENCES:  100
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SCIENCE & TECHNOLOGY LAW GROUP
;      STREET:  268 Bush Street, Suite 3200
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/121,713D
;      FILING DATE:  13-SEP-1993
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A.
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  B94-002-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)343-4341
;      TELEFAX:  (415) 343-4342
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  59:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3560 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1953
US-08-121-713D-59


  Query Match             53.7%;  Score 18.8;  DB 1;  Length 3560;
  Best Local Similarity   76.7%;  Pred. No. 34;
  Matches   23;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      2 atgtctagaatggggatggggtcactctgg 31
          || || ||| || | ||||| ||||||| |
Db   1091 ATCTCCAGATTGCGTATGGGTTCACTCTTG 1062


RESULT  11
US-08-835-268-59/c
; Sequence 59, Application US/08835268
; Patent No. 5807826
;  GENERAL INFORMATION:
;    APPLICANT:  Goodman, Corey S.
;    APPLICANT:  Kolodkin, Alex L.
;    APPLICANT:  Matthes, David
;    APPLICANT:  Bentley, David R.
;    APPLICANT:  O'Connor, Timothy
;    TITLE OF INVENTION:  The Semaphorin Gene Family
;    NUMBER OF SEQUENCES:  100
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SCIENCE & TECHNOLOGY LAW GROUP
;      STREET:  268 Bush Street, Suite 3200
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/835,268
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/121,713
;      FILING DATE:  13-SEP-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A.
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  B94-002-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)343-4341
;      TELEFAX:  (415) 343-4342
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  59:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3560 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1953
US-08-835-268-59


  Query Match             53.7%;  Score 18.8;  DB 1;  Length 3560;
  Best Local Similarity   76.7%;  Pred. No. 34;
  Matches   23;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      2 atgtctagaatggggatggggtcactctgg 31
          || || ||| || | ||||| ||||||| |
Db   1091 ATCTCCAGATTGCGTATGGGTTCACTCTTG 1062


RESULT  12
US-09-060-692-59/c
; Sequence 59, Application US/09060692
; Patent No. 5935865
;  GENERAL INFORMATION:
;    APPLICANT:  Goodman, Corey S.
;    APPLICANT:  Kolodkin, Alex L.
;    APPLICANT:  Matthes, David
;    APPLICANT:  Bentley, David R.
;    APPLICANT:  O'Connor, Timothy
;    TITLE OF INVENTION:  The Semaphorin Gene Family
;    NUMBER OF SEQUENCES:  100
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SCIENCE & TECHNOLOGY LAW GROUP
;      STREET:  268 Bush Street, Suite 3200
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/060,692
;      FILING DATE:
```

```
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/121,713
;      FILING DATE:  13-SEP-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A.
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  B94-002-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)343-4341
;      TELEFAX:  (415) 343-4342
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  59:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3560 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1953
US-09-060-692-59


  Query Match             53.7%;  Score 18.8;  DB 2;  Length 3560;
  Best Local Similarity   76.7%;  Pred. No. 34;
  Matches   23;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      2 atgtctagaatggggatggggtcactctgg 31
          || || ||| || | ||||| ||||||| |
Db   1091 ATCTCCAGATTGCGTATGGGTTCACTCTTG 1062


RESULT  13
US-08-833-391-59/c
; Sequence 59, Application US/08833391
; Patent No. 6013781
;  GENERAL INFORMATION:
;    APPLICANT:  Goodman, Corey S.
;    APPLICANT:  Kolodkin, Alex L.
;    APPLICANT:  Matthes, David
;    APPLICANT:  Bentley, David R.
;    APPLICANT:  O'Connor, Timothy
;    TITLE OF INVENTION:  The Semaphorin Gene Family
;    NUMBER OF SEQUENCES:  100
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SCIENCE & TECHNOLOGY LAW GROUP
;      STREET:  268 Bush Street, Suite 3200
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/833,391
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/121,713
;      FILING DATE:  13-SEP-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A.
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  B94-002-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)343-4341
;      TELEFAX:  (415) 343-4342
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  59:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3560 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1953
US-08-833-391-59


  Query Match             53.7%;  Score 18.8;  DB 3;  Length 3560;
  Best Local Similarity   76.7%;  Pred. No. 34;
  Matches   23;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      2 atgtctagaatggggatggggtcactctgg 31
          || || ||| || | ||||| ||||||| |
Db   1091 ATCTCCAGATTGCGTATGGGTTCACTCTTG 1062


RESULT  14
US-09-060-610-59/c
; Sequence 59, Application US/09060610
; Patent No. 6344544
;  GENERAL INFORMATION:
;    APPLICANT:  Goodman, Corey S.
;    APPLICANT:  Kolodkin, Alex L.
;    APPLICANT:  Matthes, David
;    APPLICANT:  Bentley, David R.
;    APPLICANT:  O'Connor, Timothy
;    TITLE OF INVENTION:  The Semaphorin Gene Family
;    NUMBER OF SEQUENCES:  100
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SCIENCE & TECHNOLOGY LAW GROUP
;      STREET:  268 Bush Street, Suite 3200
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/060,610
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/835,268
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A.
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  B94-002-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)343-4341
;      TELEFAX:  (415) 343-4342
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  59:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3560 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1953
US-09-060-610-59
```

```
  Query Match             53.7%;  Score 18.8;  DB 4;  Length 3560;
  Best Local Similarity   76.7%;  Pred. No. 34;
  Matches   23;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       2 atgtctagaatggggatggggtcactctgg 31
           || || ||| || | ||||| ||||||| |
Db    1091 ATCTCCAGATTGCGTATGGGTTCACTCTTG 1062


RESULT  15
PCT-US94-10151A-59/c
; Sequence 59, Application PC/TUS9410151A
;  GENERAL INFORMATION:
;    APPLICANT:  The Regents of the University of California
;    TITLE OF INVENTION:  The Semaphorin Gene Family
;    NUMBER OF SEQUENCES:  66
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR HOHBACH TEST ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/10151A
;      FILING DATE:  13-SEP-1994
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A.
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  FP-58750-PC/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299 FHT UR
;  INFORMATION FOR SEQ ID NO:  59:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3560 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1953
PCT-US94-10151A-59


  Query Match             53.7%;  Score 18.8;  DB 5;  Length 3560;
  Best Local Similarity   76.7%;  Pred. No. 34;
  Matches   23;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       2 atgtctagaatggggatggggtcactctgg 31
           || || ||| || | ||||| ||||||| |
Db    1091 ATCTCCAGATTGCGTATGGGTTCACTCTTG 1062


Search completed: June 20, 2002, 14:28:44
Job time: 18647 sec
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2002, 12:17:01 ; Search time 4682.41 Seconds
(without alignments)
100.887 Million cell updates/sec

Title: US-09-927-267-13
Perfect score: 35
Sequence: 1 catgtctagaatggggatggggtcactctggacct 35

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	24.4	69.7	938	10	BM473083	BM473083 AGENCOURT
c	2	23.6	67.4	325	9	BB272247	BB272247 BB272247
c	3	23.6	67.4	341	9	BB251608	BB251608 BB251608
c	4	23.2	66.3	288	9	BB379734	BB379734 BB379734
c	5	22.2	63.4	277	9	AW242910	AW242910 xn27e08.x
c	6	22	62.9	237	9	BB184543	BB184543 BB184543
c	7	22	62.9	285	9	BB193961	BB193961 BB193961
c	8	22	62.9	311	9	BB254958	BB254958 BB254958
c	9	22	62.9	312	9	BB275670	BB275670 BB275670
c	10	21.8	62.3	486	9	AW008857	AW008857 ws73h05.x
c	11	21.6	61.7	323	9	BB249757	BB249757 BB249757
	12	21.4	61.1	386	12	AZ831638	AZ831638 2M0111B03
	13	21.4	61.1	405	9	BB818322	BB818322 BB818322
	14	21.4	61.1	912	10	BF123776	BF123776 601760030
	15	21.2	60.6	648	12	AZ952628	AZ952628 2M0217H06
c	16	21	60.0	574	12	AZ605549	AZ605549 1M0426L21
	17	21	60.0	659	9	BB621463	BB621463 BB621463
	18	20.8	59.4	362	10	D48062	D48062 RICS13994A
c	19	20.8	59.4	436	12	FR0002586	Z86369 F.rubripes
	20	20.8	59.4	554	9	AI687518	AI687518 tp88d02.x
	21	20.8	59.4	732	12	AZ876532	AZ876532 2M0191J05
	22	20.8	59.4	834	10	BG822972	BG822972 602727981
c	23	20.6	58.9	321	10	BI189854	BI189854 g2c07fs.r
	24	20.6	58.9	393	12	AQ445686	AQ445686 nbxb0052B
	25	20.6	58.9	439	10	BF949914	BF949914 MR3-NN021
	26	20.6	58.9	448	12	AQ492113	AQ492113 V113B11 m
	27	20.6	58.9	701	12	AG067034	AG067034 Pan trogl
	28	20.6	58.9	946	10	BG683988	BG683988 602651491
c	29	20.6	58.9	1046	10	BM456459	BM456459 AGENCOURT
c	30	20.4	58.3	243	9	BB424145	BB424145 BB424145
c	31	20.4	58.3	249	10	D56898	D56898 HUM235B06B
c	32	20.4	58.3	275	9	BB190201	BB190201 BB190201
c	33	20.4	58.3	286	9	BB245918	BB245918 BB245918
	34	20.4	58.3	286	12	BH221062	BH221062 1006099B0
c	35	20.4	58.3	295	9	BB260045	BB260045 BB260045
c	36	20.4	58.3	313	9	BB250006	BB250006 BB250006
	37	20.4	58.3	398	9	AI851080	AI851080 UI-M-BH0-
c	38	20.4	58.3	400	9	BB751199	BB751199 BB751199
c	39	20.4	58.3	424	9	AA271195	AA271195 vb79c10.r
	40	20.4	58.3	450	12	AZ269508	AZ269508 RPCI-23-1
c	41	20.4	58.3	474	9	AA016579	AA016579 mg89e01.r
c	42	20.4	58.3	478	10	W82389	W82389 me80c01.r1
c	43	20.4	58.3	500	10	BG267153	BG267153 1000112B0
	44	20.4	58.3	534	9	AW747153	AW747153 WS1_66_G0
	45	20.4	58.3	534	10	BE597022	BE597022 PI1_60_D0

ALIGNMENTS

RESULT 1
BM473083/c
LOCUS BM473083 938 bp mRNA linear EST 05-FEB-2002
DEFINITION AGENCOURT_6466078 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5574682 5', mRNA sequence.
ACCESSION BM473083
VERSION BM473083.1 GI:18522125
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 938)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12323 row: 1 column: 11
High quality sequence start: 7
High quality sequence stop: 706.
FEATURES Location/Qualifiers
source 1. .938
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5574682"
/clone_lib="NIH_MGC_88"
/tissue_type="duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: small intestine; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.767 kb. Library enriched for full-length clones and constructed by Life

Technologies. Note: this is a NIH_MGC Library."
BASE COUNT 271 a 252 c 193 g 221 t 1 others
ORIGIN

Query Match 69.7%; Score 24.4; DB 10; Length 938;
Best Local Similarity 96.2%; Pred. No. 40;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 9 gaatggggatggggtcactctggacc 34
| ||||||||||||||||||||||||||
Db 63 GGATGGGGATGGGGTCACTCTGGACC 38

RESULT 2
BB272247/c
LOCUS BB272247 325 bp mRNA linear EST 07-JUL-2000
DEFINITION BB272247 RIKEN full-length enriched, 10 days neonate cortex Mus musculus cDNA clone A830067B17 3', mRNA sequence.
ACCESSION BB272247
VERSION BB272247.1 GI:8969266
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 325)
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M. , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y. , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y. , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I. , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino ,M., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL Unpublished (2000)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki ,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1. .325
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A830067B17"
/clone_lib="RIKEN full-length enriched, 10 days neonate cortex"
/tissue_type="cortex"
/dev_stage="10 days neonate"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT 96 a 77 c 49 g 103 t
ORIGIN

Query Match 67.4%; Score 23.6; DB 9; Length 325;
Best Local Similarity 86.7%; Pred. No. 67;
Matches 26; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 catgtctagaatggggatggggtcactctg 30
| ||||||||| ||||||||| |||||| |
Db 144 CTTGTCTAGAAAGGGGATGGGATCACTCAG 115

RESULT 3
BB251608/c
LOCUS BB251608 341 bp mRNA linear EST 06-JUL-2000
DEFINITION BB251608 RIKEN full-length enriched, 7 days neonate cerebellum Mus musculus cDNA clone A730047E09 3', mRNA sequence.
ACCESSION BB251608
VERSION BB251608.1 GI:8944354
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 341)
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M. , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y. , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y. , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I. , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino ,M., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL Unpublished (2000)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)

Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki ,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

FEATURES Location/Qualifiers
source 1. .341
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A730047E09"
/clone_lib="RIKEN full-length enriched, 7 days neonate cerebellum"
/tissue_type="cerebellum"
/dev_stage="7 days neonate"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 97 a 75 c 61 g 108 t
ORIGIN

Query Match 67.4%; Score 23.6; DB 9; Length 341;
Best Local Similarity 86.7%; Pred. No. 67;
Matches 26; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy      1 catgtctagaatggggatggggtcactctg 30
          | |||||||||| |||||||| |||||| |
Db    161 CTTGTCTAGAATTGGGATGGGATCACTCAG 132
```

RESULT 4
BB379734/c
LOCUS BB379734 288 bp mRNA linear EST 13-JUL-2000
DEFINITION BB379734 RIKEN full-length enriched, 0 day neonate cerebellum Mus musculus cDNA clone C230001G20 3', mRNA sequence.
ACCESSION BB379734
VERSION BB379734.1 GI:9092624
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 288)
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M. , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y. , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y. , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I. , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino ,M., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL Unpublished (2000)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki ,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

FEATURES Location/Qualifiers
source 1. .288
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="C230001G20"
/clone_lib="RIKEN full-length enriched, 0 day neonate cerebellum"
/tissue_type="cerebellum"
/dev_stage="0 day neonate"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 479.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 81 a 71 c 39 g 97 t
ORIGIN

Query Match 66.3%; Score 23.2; DB 9; Length 288;
Best Local Similarity 89.3%; Pred. No. 92;
Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy      3 tgtctagaatggggatggggtcactctg 30
          ||| |||||||||||||| |||||| |
Db    106 TGTGTAGAATGGGGATGGGATCACTCAG 79
```

RESULT 5
AW242910/c
LOCUS AW242910 277 bp mRNA linear EST 14-DEC-1999
DEFINITION xn27e08.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2694950 3' similar to TR:Q61204 Q61204 NOTCH2-LIKE ;, mRNA sequence.

```
ACCESSION   AW242910
VERSION     AW242910.1  GI:6576679
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 277)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40UP from Gibco
            High quality sequence stop: 147.
FEATURES             Location/Qualifiers
     source          1. .277
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2694950"
                     /clone_lib="NCI_CGAP_Kid11"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
                     prepared, and ss circles were made in vitro. Following HAP
                     purification, this DNA was used as tracer in a subtractive
                     hybridization reaction. The driver was PCR-amplified cDNAs
                     from a pool of 5,000 clones made from the same library
                     (cloneIDs 1322376-1323911, 1456007-1456775, and
                     1500552-1502855). Subtraction by Bento Soares and M.
                     Fatima Bonaldo. "
BASE COUNT       89 a     48 c     46 g     94 t
ORIGIN


  Query Match             63.4%;  Score 22.2;  DB 9;  Length 277;
  Best Local Similarity   77.1%;  Pred. No. 2.2e+02;
  Matches   27;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      1 catgtctagaatggggatggggtcactctggacct 35
          | ||| | |||||||| | ||||||| |||| | ||
Db    248 CTTGTTTTGAATGGGGTTTGGGTCAGTCTGCAGCT 214


RESULT   6
BB184543/c
LOCUS       BB184543                237 bp    mRNA    linear   EST 29-JUN-2000
DEFINITION  BB184543 RIKEN full-length enriched, adult male spinal cord Mus
            musculus cDNA clone A330012012 3', mRNA sequence.
ACCESSION   BB184543
VERSION     BB184543.1  GI:8843626
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 237)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .237
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="A330012012"
                     /clone_lib="RIKEN full-length enriched, adult male spinal
                     cord"
                     /sex="male"
                     /tissue_type="spinal cord"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 459.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC
                     3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I. "
BASE COUNT       73 a     45 c     36 g     83 t
ORIGIN


  Query Match             62.9%;  Score 22;  DB 9;  Length 237;
  Best Local Similarity   83.3%;  Pred. No. 2.5e+02;
  Matches   25;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 catgtctagaatggggatggggtcactctg 30
          | |||||||||| || ||||| |||||| |
Db     57 CTTGTCTAGAATTGGCATGGGATCACTCAG 28
```

```
RESULT   7
BB193961/c
LOCUS       BB193961                285 bp    mRNA    linear   EST 30-JUN-2000
DEFINITION  BB193961 RIKEN full-length enriched, adult male spinal cord Mus
            musculus cDNA clone A330080B13 3', mRNA sequence.
ACCESSION   BB193961
VERSION     BB193961.1  GI:8854580
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 285)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .285
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="A330080B13"
                     /clone_lib="RIKEN full-length enriched, adult male spinal
                     cord"
                     /sex="male"
                     /tissue_type="spinal cord"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 459.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I. "
BASE COUNT       82 a     53 c     54 g     96 t
ORIGIN


  Query Match             62.9%;  Score 22;  DB 9;  Length 285;
  Best Local Similarity   83.3%;  Pred. No. 2.6e+02;
  Matches   25;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 catgtctagaatggggatggggtcactctg 30
           | |||||||||| |||||||| ||||| |
Db     104 CTTGTCTAGAATTGGGATGGGACCACTCAG 75


RESULT   8
BB254958/c
LOCUS       BB254958                311 bp    mRNA    linear   EST 06-JUL-2000
DEFINITION  BB254958 RIKEN full-length enriched, 7 days neonate cerebellum Mus
            musculus cDNA clone A730064I15 3', mRNA sequence.
ACCESSION   BB254958
VERSION     BB254958.1  GI:8947704
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 311)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .311
                     /organism="Mus musculus"
```

```
                     /db_xref="taxon:10090"
                     /clone="A730064I15"
                     /clone_lib="RIKEN full-length enriched, 7 days neonate
                     cerebellum"
                     /tissue_type="cerebellum"
                     /dev_stage="7 days neonate"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 20.0 and subtraction to Rot = 459.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC
                     3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I."
BASE COUNT       94 a     58 c     58 g    101 t
ORIGIN


  Query Match             62.9%;  Score 22;  DB 9;  Length 311;
  Best Local Similarity   83.3%;  Pred. No. 2.6e+02;
  Matches   25;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 catgtctagaatggggatggggtcactctg 30
          | |||||||||| |||||||  |||||| |
Db    133 CTTGTCTAGAATTGGGATGGCATCACTCAG 104


RESULT  9
BB275670/c
LOCUS       BB275670                312 bp    mRNA    linear   EST 07-JUL-2000
DEFINITION  BB275670 RIKEN full-length enriched, 10 days neonate cortex Mus
            musculus cDNA clone A830094B15 3', mRNA sequence.
ACCESSION   BB275670
VERSION     BB275670.1  GI:8972691
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 312)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .312
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="A830094B15"
                     /clone_lib="RIKEN full-length enriched, 10 days neonate
                     cortex"
                     /tissue_type="cortex"
                     /dev_stage="10 days neonate"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 20.0 and subtraction to Rot = 459.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC
                     3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I."
BASE COUNT       90 a     54 c     59 g    109 t
ORIGIN


  Query Match             62.9%;  Score 22;  DB 9;  Length 312;
  Best Local Similarity   83.3%;  Pred. No. 2.6e+02;
  Matches   25;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 catgtctagaatggggatggggtcactctg 30
          | |||||||||| |||||||  |||||| |
Db    131 CTTGTCTAGAATTGGGATGGCATCACTCAG 102


RESULT  10
AW008857/c
LOCUS       AW008857                486 bp    mRNA    linear   EST 08-MAR-2000
DEFINITION  ws73h05.x1 NCI_CGAP_Co3 Homo sapiens cDNA clone IMAGE:2503641 3'
            similar to TR:Q61204 Q61204 NOTCH2-LIKE ;, mRNA sequence.
ACCESSION   AW008857
VERSION     AW008857.1  GI:5857635
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 486)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
```

Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 680 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 314.

FEATURES Location/Qualifiers
source 1. .486
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2503641"
/clone_lib="NCI_CGAP_Co3"
/sex="pooled"
/tissue_type="colon"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from 12 pooled bulk tumor samples and primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. "

BASE COUNT 131 a 100 c 94 g 161 t
ORIGIN

Query Match 62.3%; Score 21.8; DB 9; Length 486;
Best Local Similarity 78.8%; Pred. No. 3.3e+02;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy      3 tgtctagaatggggatggggtcactctggacct 35
          ||| | |||||||| | |||||| |||| | ||
Db    250 TGTTTTGAATGGGGTTTGGGTCAGTCTGCAGCT 218
```

RESULT 11
BB249757/c
LOCUS BB249757 323 bp mRNA linear EST 06-JUL-2000
DEFINITION BB249757 RIKEN full-length enriched, 7 days neonate cerebellum Mus musculus cDNA clone A730035D17 3', mRNA sequence.
ACCESSION BB249757
VERSION BB249757.1 GI:8942503
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 323)
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I., Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL Unpublished (2000)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

FEATURES Location/Qualifiers
source 1. .323
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A730035D17"
/clone_lib="RIKEN full-length enriched, 7 days neonate cerebellum"
/tissue_type="cerebellum"
/dev_stage="7 days neonate"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 87 a 76 c 41 g 118 t 1 others
ORIGIN

Query Match 61.7%; Score 21.6; DB 9; Length 323;
Best Local Similarity 85.7%; Pred. No. 3.7e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy      1 catgtctagaatggggatggggtcactc 28
          | |||||| |||||||||||  ||||||
Db    143 CTTGTCTAAAATGGGGATGGCATCACTC 116
```

RESULT 12
AZ831638
LOCUS AZ831638 386 bp DNA linear GSS 20-FEB-2001
DEFINITION 2M0111B03R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0111B03 R, DNA sequence.
ACCESSION AZ831638
VERSION AZ831638.1 GI:13001546
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 386)

```
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0111  row: B  column: 03
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 386.
FEATURES             Location/Qualifiers
     source          1. .386
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0111B03"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      120 a     78 c    100 g     88 t
ORIGIN


  Query Match             61.1%;  Score 21.4;  DB 12;  Length 386;
  Best Local Similarity   80.6%;  Pred. No. 4.5e+02;
  Matches   25;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       5 tctagaatggggatggggtcactctggacct 35
           |||||||| |||||| |||   |||||| ||
Db     195 TCTAGAATCGGGATGAGGTTGATCTGGAACT 225


RESULT  13
BE818322
LOCUS       BB818322                405 bp    mRNA    linear   EST 19-NOV-2001
DEFINITION  BB818322 RIKEN full-length enriched, lung RCB-0558 LLC cDNA Mus
            musculus cDNA clone G730044J10 3', mRNA sequence.
ACCESSION   BB818322
VERSION     BB818322.1  GI:16990951
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 405)
  AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
            Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
            ,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
            Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.
            , Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
            Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
            ,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.
            , Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
            2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .405
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="G730044J10"
                     /clone_lib="RIKEN full-length enriched, lung RCB-0558 LLC
                     cDNA"
                     /tissue_type="lung"
                     /cell_line="RCB-0558 LLC"
BASE COUNT      119 a    106 c     84 g     96 t
ORIGIN


  Query Match             61.1%;  Score 21.4;  DB 9;  Length 405;
  Best Local Similarity   80.6%;  Pred. No. 4.5e+02;
  Matches   25;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 catgtctagaatggggatggggtcactctgg 31
           |||||||||||| ||||| |||| | || | |
Db     200 CATGTCTAGAAAGGGGAAGGGGGCCCTTTCG 230


RESULT  14
BF123776
LOCUS       BF123776                912 bp    mRNA    linear   EST 24-OCT-2000
DEFINITION  601760030F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:4023294 5',
            mRNA sequence.
ACCESSION   BF123776
VERSION     BF123776.1  GI:10962816
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 912)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9281  row: k  column: 07
            High quality sequence stop: 738.
FEATURES              Location/Qualifiers
     source           1. .912
                      /organism="Mus musculus"
                      /strain="C57/B6"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:4023294"
                      /clone_lib="NCI_CGAP_Mam5"
                      /tissue_type="tumor, gross tissue"
                      /dev_stage="7 months"
                      /lab_host="DH10B"
                      /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                      Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                      Library constructed by Life Technologies. Investigators
                      providing samples: Lothar Hennighausen/Robin Humphreys,
                      NIH"
BASE COUNT      225 a    256 c    166 g    265 t
ORIGIN


  Query Match             61.1%;  Score 21.4;  DB 10;  Length 912;
  Best Local Similarity   80.6%;  Pred. No. 5.1e+02;
  Matches   25;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 catgtctagaatggggatggggtcactctgg 31
          ||||| ||  ||||||| |||||| || |||
Db    826 CATGTTTACCATGGGGACGGGGTCTCTTTGG 856


RESULT 15
AZ952628
LOCUS       AZ952628                 648 bp    DNA     linear   GSS 27-APR-2001
DEFINITION  2M0217H06R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
            clone UUGC2M0217H06 R, DNA sequence.
ACCESSION   AZ952628
VERSION     AZ952628.1  GI:13823855
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 648)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0217  row: H  column: 06
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 648.
FEATURES              Location/Qualifiers
     source           1. .648
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="UUGC2M0217H06"
                      /clone_lib="Mouse 10kb plasmid UUGC2M library"
                      /sex="Female"
                      /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
                      /note="Vector: PWD42nv; Purified genomic DNA from M.
                      musculus C57BL/6J (female) was obtained  from the Jackson
                      Laboratory Mouse DNA Resource
                      (http://www.jax.org/resources/documents/dnares/). The DNA
                      was hydrodynamically sheared by repeated passage through a
                      0.005 inch orifice at constant velocity. The sheared DNA
                      was blunt end-repaired with T4 DNA polymerase and T4
                      polynucleotide kinase. Adaptor oligonucleotides were
                      ligated to the blunt ends in high  molar excess. The
                      adaptored DNA was purified and size-selected for a 9.5 to
                      10.5 kb range using preparative agarose gel
                      electrophoresis. Vector DNA was prepared from a derivative
                      of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                      inducible derivative of plasmid R1. The vector was ligated
                      with adaptors complementary to the insert adaptors and
                      purified. The sheared, adaptored mouse DNA was annealed to
                      adaptored vector DNA, and transformed into
                      chemically-competent E. coli XL10-Gold (Stratagene) cells
                      and selected for ampicillin resistance."
BASE COUNT      183 a    157 c    137 g    171 t
ORIGIN


  Query Match             60.6%;  Score 21.2;  DB 12;  Length 648;
  Best Local Similarity   88.5%;  Pred. No. 5.8e+02;
  Matches   23;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 atgtctagaatggggatggggtcact 27
          | ||||| ||||| ||||||||||||
Db    149 AAGTCTACAATGGTGATGGGGTCACT 174


Search completed: June 20, 2002, 12:17:05
Job time: 11378 sec
```